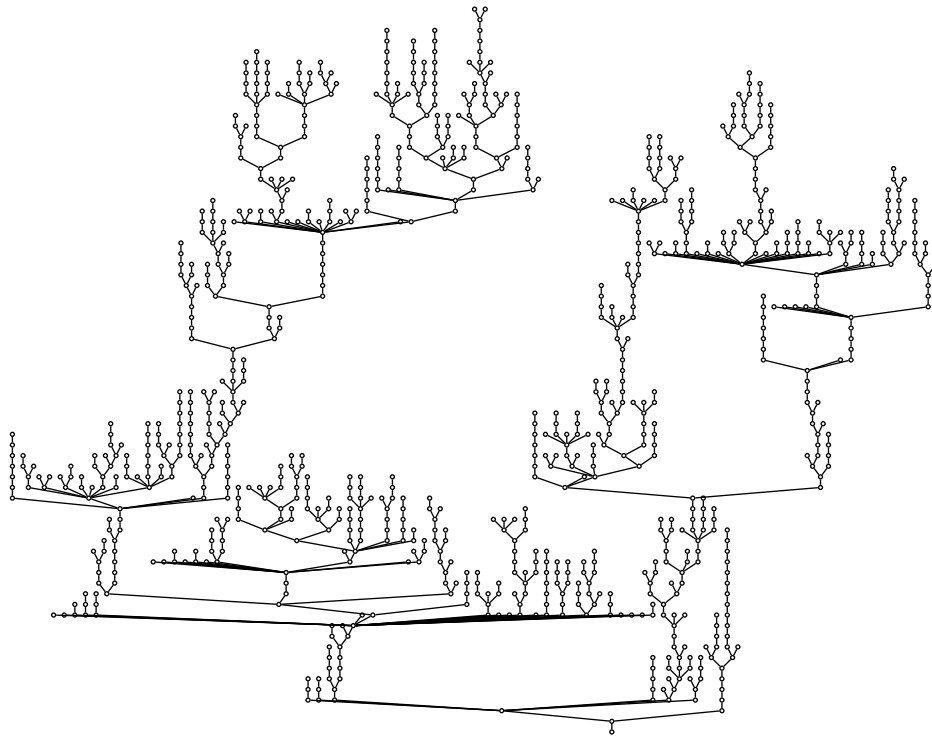


# Große Stammbäume

Dissertation  
zur Erlangung des Doktorgrades  
der Naturwissenschaften



vorgelegt beim Fachbereich Mathematik  
der Johann Wolfgang Goethe-Universität  
in Frankfurt am Main

von

Lars Kauffmann  
aus Friedberg (Hessen)

Frankfurt am Main (2003)

(DF1)

vom Fachbereich Mathematik  
der Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan :

Gutachter : Prof. Dr. G. Kersting , Prof. Dr. A.Wakolbinger

Datum der Disputation :

Sei  $T$  ein kritischer oder subkritischer Galton-Watson Stammbaum (GW-Baum) mit einer Kinderzahlverteilung endlicher oder unendlicher Varianz. Wir sind an der Struktur von  $T$ , bedingt darauf, dass  $T$  “groß” ist, interessiert. Der klassische sowie naheliegende Zugang ist,  $T$  auf eine große Gesamtgröße oder eine große Höhe zu bedingen. In dieser Arbeit werden drei, zum GW-Baum eng verwandte Typen von zufälligen Stammbäumen vorgestellt, deren Analyse aufschlussreiche Einsichten über große GW-Stammbäume liefert.

Zur Untersuchung dieser auf große Gesamtgröße bedingten Stammbäume schlagen wir eine Familie von zufälligen, größenverzerrten Bäumen vor, deren auf Größe bedingte Verteilung mit der des, auf gegebener Größe bedingten, Baumes  $T$  übereinstimmt. Diese zufälligen Stammbäume besitzen eine einfache probabilistische Struktur, wenn man sie entlang der Ahnenlinien von rein zufällig gezogenen Knoten zerlegt. Die Verwandtschaftsstruktur des von den gezogenen Knoten und der Wurzel aufgespannten Teilbaumes hängt im wesentlichen von dem asymptotischen Verhalten der Kinderzahlverteilung ab. Während bei endlicher Varianz diese Teilbäume asymptotisch binär sind, können bei unendlicher Varianz im Limes auch andere Formen auftreten. Wir zeigen, dass diese Teilbäume GW-Bäume bedingt auf ihre Gesamtblätterzahl sind. Mit Hilfe der Zerlegung entlang der Ahnenlinien erhalten wir zudem einen Grenzwertsatz für die reskalierte Gesamtgröße des Baumes mit einer Gamma-Verteilung als Limes. Die Analyse großer Bäume führen wir unter dem Aspekt des Größenverzerrrens fort, indem wir eine weitere Familie zufälliger Bäume vorschlagen. Diese erhalten wir durch Größenverzerrung in der  $n$ -ten Generationsgröße. Wir werden sehen, dass der dadurch gewonnene zufällige Stammbaum eine ähnliche probabilistische Struktur wie der in der Gesamtgröße größenverzerrte Baum besitzt. Hier beweisen wir mit einfachen Überlegungen Aussagen über die Generation des jüngsten gemeinsamen Vorfahren (MRCA) von uniform aus Generation  $n$  gezogenen Knoten, sowie die Struktur des von diesen Knoten aufgespannten Skeletts.

Schließlich betrachten wir die in [15] vorgestellte probabilistische Zerlegung des auf Mindesthöhe  $n$  bedingten GW-Baumes. Damit werden wir klassische Sätze über die Höhe des MRCA und die Grenzverteilung der reskalierten  $n$ -ten Generationsgröße für den Fall einer Kinderzahlverteilung mit unendlicher Varianz auf alternativem und anschaulichem Weg beweisen. Zudem erhalten wir eine Grenzverteilung für die Anzahl der Kinder des MRCA.

---

# Inhaltsverzeichnis

<b>1</b>	<b>Einleitung, Überblick und Notationen</b>	<b>7</b>
<b>2</b>	<b>Größenverzerrte Galton-Watson Bäume</b>	<b>17</b>
2.1	Größenverzerrung in der Gesamtgröße . . . . .	18
2.2	Konstruktion entlang eines zufälligen Skeletts . . . . .	20
2.3	Limes des reduzierten Baumes . . . . .	26
2.4	Größe des größenverzerrten Baumes . . . . .	36
2.5	Größenverzerrung in $n$ -ter Generation . . . . .	44
<b>3</b>	<b>Galton-Watson Bäume bedingt auf Mindesthöhe</b>	<b>51</b>
3.1	Konstruktion des bedingten GW-Baumes . . . . .	52
3.2	Grenzwertsätze über den bedingten GW-Baum . . . . .	56
	<b>Literaturverzeichnis</b>	<b>69</b>



# Kapitel 1

## Einleitung, Überblick und Notationen

Wir betrachten einen Galton-Watson Verzweigungsprozess mit Kinderzahlverteilung  $(p_j)_{j \geq 0}$ . Dieser Verzweigungsprozess wird durch einen lokalen Verzweigungsmechanismus beschrieben: Jedes Individuum besitzt unabhängig von den anderen Individuen Kinder gemäß der Verteilung  $(p_j)_{j \geq 0}$ , d.h. mit Wahrscheinlichkeit  $p_j$  hat ein Individuum  $j$  Kinder. Man erhält den Stammbaum  $T$  des Verzweigungsprozesses, indem man jedes Individuum durch einen Knoten repräsentiert und jeden Knoten mit seinen Kindern durch jeweils eine Kante verbindet. Wir fassen dabei  $T$  als einen verwurzelten, planaren Baum auf, in dem die unterscheidbaren Kinder jedes Knotens von links nach rechts angeordnet sind. Die Wurzel des Baumes ist dabei der einzige Knoten ohne Vorfahr.

Das wohl bekannteste Theorem über GW-Prozesse ist das Theorem über die Größe kritischer GW-Bäume, welches besagt, dass die Gesamtpopulation (=Anzahl der Knoten) fast sicher endlich ist, genau dann, wenn die erwartete Anzahl der Nachkommen pro Knoten höchstens 1 ist:

$$\mathbb{P}(T \text{ endlich}) = 1 \iff \mu := \sum_{j=0}^{\infty} j p_j \leq 1 \text{ und } p_1 < 1. \quad (1.1)$$

Um nun die Struktur eines “großen” *kritischen* oder *subkritischen* Galton-Watson Baumes  $T$  (d.h.  $\mu \leq 1$ ) zu untersuchen, hat man im wesentlichen zwei Möglichkeiten die Anzahl der Individuen des Baumes groß werden zu lassen. Man bedingt den Baum  $T$  auf exakte (oder Mindest-) Höhe  $n$  oder auf exakt (oder

mindestens)  $n$  Knoten im gesamten Baum. Im kritischen Fall ( $\mu = 1$ ) sind diese beiden bedingten Verteilungen asymptotisch gleich – im subkritischen Fall hingegen essentiell verschieden (vgl. [4]). Ein ähnlicher und verwandter Effekt kennt man für Irrfahrten bedingt darauf, im Positiven zu sein (siehe [7]). Die Struktur des auf Höhe bedingten GW-Baumes kennt man schon bestens (siehe [15, 17]). Wir werden diese in Kapitel 3 verwenden, um klassische Resultate auf alternativem und aufschlussreichem Weg zu beweisen. Der auf Gesamtgröße bedingte Baum mit Kinderzahlverteilung endlicher Varianz wurde in [22] und in der kombinatorischen Literatur untersucht (siehe z.B. [24, 26]). Im kombinatorischen Kontext entspricht der bedingte Baum dem einfach erzeugten Baum (*simply generated tree*). Grundlagen über Galton-Watson Prozesse findet man z.B. in [5].

In Abschnitt 2.1 dieser Arbeit schlagen wir einen neuen Zugang vor. Dieser besitzt eine enge Beziehung zum Bedingen des Baumes auf Gesamtgröße. Wir untersuchen eine Familie von Verteilungen  $\mathcal{L}_{\theta,k}$ ,  $0 < \theta < 1$ ,  $k \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$  eines zufälligen Stammbaumes  $\hat{T}$  mit den folgenden Eigenschaften:

1. Für jedes  $\theta$  und  $k$  stimmt die auf Gesamtgröße bedingte Verteilung des zufälligen Baumes  $\hat{T}$  mit der des auf Gesamtgröße bedingten GW-Baumes  $T$  überein, d.h.

$$\mathcal{L}_{\theta,k}(\hat{T} \mid s(\hat{T}) = n) = \mathcal{L}(T \mid s(T) = n), \quad n \geq 1, \quad (1.2)$$

wobei  $s(T)$  die Größe (= Anzahl der Knoten) des Baumes  $T$  und  $\mathcal{L}$  die Verteilung eines GW-Baumes  $T$  mit Kinderzahlverteilung  $(p_j)_{j \geq 0}$  bezeichnet.

2. Sei  $h(\theta, k)$  die erwartete Größe des Baumes  $\hat{T}$  unter  $\mathcal{L}_{\theta,k}$ ,

$$h(\theta, k) := \mathbb{E}_{\theta,k} s(\hat{T}). \quad (1.3)$$

Für alle  $\theta$  wächst die Funktion  $h$  in  $k$  und für alle  $k$  gilt

$$h(\theta, k) \uparrow \infty \quad \text{für } \theta \uparrow 1. \quad (1.4)$$

Der Parameter  $\theta$  kann dazu benutzt werden, die erwartete Größe des Baumes einzustellen. Wir werden sehen, dass die Verteilung der Gesamtgröße  $\mathcal{L}_{\theta,k}(s(\hat{T}))$  sich im Limes  $\theta \uparrow 1$  nicht um den Erwartungswert konzentriert, können aber bei



passender Reskalierung von  $s(\hat{T})$  unter  $\mathcal{L}_{\theta,k}$  eine Gamma Grenzverteilung herleiten (Theorem 2.8). Im Falle des auf Gesamtgröße bedingten Baumes verliert man die Unabhängigkeit der Kinderzahlen der einzelnen Individuen. In unserem Ansatz mit zufälliger Größe hingegen, erhalten wir eine interessante und aufschlussreiche Unabhängigkeitsstruktur des Stammbaumes  $\hat{T}$ .

Die Bedeutung des Parameters  $k$  ist wie folgt: Zur Untersuchung der Struktur großer Bäume ist es üblich, den Teilbaum aufgespannt von einer Anzahl rein zufällig gezogener Knoten und der Wurzel zu betrachten. Solche Teilbäume liefern ein gutes Verständnis über die “Geometrie” des Baumes. Die Verteilung  $\mathcal{L}_{\theta,k}$  passt genau in diesen Kontext, Teilbäume (aufgespannt von  $k$  rein zufällig gezogenen Knoten) von großen GW-Bäumen zu untersuchen. GW-Bäume bedingt auf Gesamtgröße und deren Teilbäume wurden von Aldous in [3] (unter der Annahme einer Kinderzahlverteilung mit endlicher Varianz) und von Duquesne und Le Gall in [10, 11] untersucht. Unser Ansatz erlaubt es auch, Kinderzahlverteilungen mit unendlicher Varianz (unter gewissen natürlichen Regularitätsbedingungen) zu behandeln. Im Fall endlicher Varianz stellt sich heraus, dass die Teilbäume asymptotisch binär sind, während im Fall unendlicher Varianz auch andere Formen den Limes mit positiver Wahrscheinlichkeit bestehen (Theorem 2.4). Genauer werden wir zeigen, dass der von  $k$  uniform aus  $\hat{T}$  gezogenen Knoten aufgespannte Teilbaum unter  $\mathcal{L}_{\theta,k}$  gegen einen, auf  $k$  Blätter bedingten, GW-Baum mit einer ausgezeichneten Kinderzahlverteilung konvergiert. Der Schlüssel zu diesem Resultat ist die Konstruktion eines zufälligen Baumes der die Verteilung  $\mathcal{L}_{\theta,k}$  besitzt.

In Abschnitt 2.5 betrachten wir, motiviert durch das Konzept der Größenverzerrung, eine Familie von Verteilungen  $\mathcal{L}_{k,n}$ ,  $k, n \in \mathbb{N}$  von zufälligen Stammbäumen  $\hat{T}$ , die wir durch Größenverzerrung in der  $n$ -ten Generation erhalten. Auch dieser Ansatz erlaubt eine mehrstufige Konstruktion eines zufälligen Baumes mit Verteilung  $\mathcal{L}_{k,n}$ . Die Konstruktion ist verwandt mit der Zerlegung von umgewichteten GW-Bäumen entlang einer Ahnenlinie (siehe [25]). Wir werden Aussagen über die Generation des MRCA von  $k$ , rein zufällig aus Generation  $n$  gezogenen Knoten beweisen, sowie den von diesen Knoten aufgespannten Teilbaum im Limes  $n \rightarrow \infty$  untersuchen. Besitzt die Kinderzahlverteilung endliche  $k$ -te Momente, so ist dieser Teilbaum im Grenzwert ein binärer Baum. Die Grenzverteilung der reskalierten Höhe des MRCA ist eine Beta-Verteilung zu den Parametern  $(1, k - 1)$ .

In Kapitel 3 beweisen wir mittels der Konstruktion des auf Höhe bedingten GW-Baumes (siehe [15, 17]) klassische Sätze über die Verteilung der Höhe des MRCA aller Knoten  $n$ -ter Generation sowie der reskalierten  $n$ -ten Generationsgröße. Wir interessieren uns dabei hauptsächlich für alternative Beweise im Fall unendlicher Varianz. Zudem erhalten wir eine Grenzverteilung für die Anzahl der Kinder des MRCA.

Für weitere Zerlegungen und verwandte Konstruktionen in verschiedenen Verzweigungskontexten siehe [9, 20, 21].

Ich bedanke mich bei Jochen Geiger und Götz Kersting für die fachliche Unterstützung und fruchtbaren Diskussionen.

Wir beginnen mit einer formalen Beschreibung von Stammbäumen und den im Verlauf der Arbeit verwendeten Begriffen.

## Formale Beschreibung von Stammbäumen, Teilbäumen und Ahnenlinien

Die folgende formale Beschreibung von Bäumen, Teilbäumen und die in diesem Kontext benötigten Begriffe und Notationen sind rein mengentheoretisch und werden nun zunächst unabhängig vom Wahrscheinlichkeitsmodell eingeführt.

Zur Beschreibung von verwurzelten, geordneten Bäumen benutzen wir die übliche Notation (siehe z.B. [19, 28]) einen Baum mit der Menge seiner Knoten zu identifizieren: Die Wurzel erhält das Symbol  $\Delta$ . Die unterscheidbaren Kinder der Wurzel der ersten Generation werden mit  $(1), (2), (3), \dots$  beschriftet. Ein Knoten in der  $m$ -ten Generation (der Abstand von der Wurzel beträgt  $m$  Kanten) notiert man durch ein  $m$ -Tupel  $(v_1, \dots, v_m)$ ,  $v_i \in \mathbb{N}$ . Dieser Knoten ist das  $v_m$ -te Kind des Knotens mit der Beschriftung  $(v_1, \dots, v_{m-1})$ . Mit  $|v|$  bezeichnen wir die Generation (oder Höhe) des Knotens  $v$ , das heißt  $v$  ist in der  $m$ -ten Generation ( $|v| = m$ ) wenn  $v \in \mathbb{N}^m$ ,  $m \geq 1$ . Wir setzen  $|\Delta| = 0$ . Für einen Knoten  $v = (v_1, \dots, v_m)$  definieren wir  $[v]_\ell := (v_1, \dots, v_\ell)$ ,  $1 \leq \ell \leq m$  und  $[v]_0 := \Delta$ . Der Knoten  $[v]_\ell$  ist der Vorfahr von  $v$  in Generation  $\ell$ . Die Knotenfolge  $([v]_\ell)_{0 \leq \ell \leq m}$  bezeichnet die *Ahnenlinie* von  $v$ .

Eine nichtleere Teilmenge  $t \subset \mathcal{V} := \{\Delta\} \cup \bigcup_{m=1}^{\infty} \mathbb{N}^m$  nennt man *verwurzelter, geordneter Baum*, falls

$$\begin{aligned} v \in t &\implies [v]_\ell \in t, \text{ für } 0 \leq \ell \leq m := |v| \text{ und} \\ &([v]_{m-1}, j) \in t, \text{ für } 1 \leq j \leq v_m. \end{aligned}$$

Dabei schreiben wir  $(v, j) \in \mathbb{N}^{|v|+1}$  für den Knoten, den man durch Anhängen des Kindes  $j \in \mathbb{N}$  an  $v \in \mathcal{V}$  erhält.

Sei  $c(v, t)$  die Anzahl der Kinder (direkten Nachfolger) von  $v$  in  $t$ , d.h.,

$$c(v, t) := |\{j \in \mathbb{N} : (v, j) \in t\}|. \quad (1.5)$$

Die Anzahl aller Knoten in  $t$  mit  $i$  Kindern notieren wir durch

$$\beta_i(t) := |\{v \in t : c(v, t) = i\}|, \quad i \geq 0.$$

Ein Knoten  $v$  in  $t$  wird *Blatt* genannt falls er keine Kinder hat, d.h. falls  $c(v, t) = 0$ . Sei  $\Lambda(t)$  die Menge der Blätter des Baumes  $t$ ,

$$\Lambda(t) = \{v \in t : c(v, t) = 0\}.$$

Für einen Knoten  $w \in t$  definieren wir den in  $w$  verwurzelten Teilbaum  $t_w$  aus  $t$  durch

$$t_w := \{v \in \mathcal{V} : (w, v) \in t\}.$$

Die Menge der Knoten der  $m$ -ten Generation von  $t$  bezeichnen wir mit

$$z_m(t) := \{v \in t : |v| = m\} \quad (1.6)$$

Wir halten hier zur späteren Verwendung noch folgende einfache Identitäten fest. Für einen endlichen Baum  $t$  gilt für die Gesamtzahl der Knoten

$$s(t) = \sum_{i \geq 0} \beta_i(t) = 1 + \sum_{i \geq 1} i \beta_i(t) \quad (1.7)$$

und folglich für die Zahl der Blätter

$$\beta_0(t) = s(t) - \sum_{i \geq 1} \beta_i(t) = 1 + \sum_{i \geq 1} (i-1) \beta_i(t). \quad (1.8)$$

Der jüngste gemeinsame Vorfahre (MRCA = Most Recent Common Ancestor) zweier Knoten  $v, w \in \mathcal{V}$  ist der Knoten

$$v \wedge w := [v]_{\max\{\ell \geq 0 : [v]_\ell = [w]_\ell\}} = [v]_{|v \wedge w|}.$$

Weiterhin definieren wir den MRCA der Knotenmenge  $M := \{v^{(1)}, \dots, v^{(k)}\}$  durch

$$\wedge(M) := v^{(1)} \wedge \dots \wedge v^{(k)} := ((\dots ((v^{(1)} \wedge v^{(2)}) \wedge v^{(3)}) \dots) \wedge v^{(k)}). \quad (1.9)$$

Offensichtlich hängt der MRCA  $v^{(1)} \wedge \dots \wedge v^{(k)}$  nicht von der Reihenfolge der Knoten  $v^{(i)}, 1 \leq i \leq k$  ab.

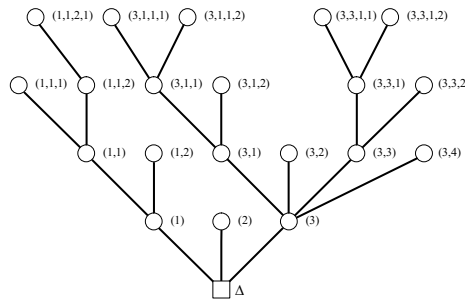


ABBILDUNG 1.1: *Beispiel eines verwurzelten, beschrifteten, geordneten Baumes*

## Skelette und reduzierte Skelette

Zur Beschreibung des Teilbaumes, der durch  $k$  Knoten eines Baumes und der Wurzel  $\Delta$  aufgespannt wird, definieren wir nun den Begriff des Skeletts:

Eine nichtleere Teilmenge  $r \subset \mathcal{V}$  heißt *verwurzeltes, geordnetes Skelett*, falls

$$v \in r \implies [v]_\ell \in r, \quad 0 \leq \ell \leq |v|.$$

Im Gegensatz zu einem Baum müssen im Skelett die jüngeren Geschwister der Knoten nicht enthalten sein. Analog (1.5) schreiben wir  $c(v, r)$  für die Anzahl der Kinder von  $v$  im Skelett  $r$ ,

$$c(v, r) := |\{j \geq 1 : (v, j) \in r\}|.$$

Mit  $r(v^{(1)}, \dots, v^{(k)})$  bezeichnen wir das Skelett welches durch  $k$  verschiedene Knoten  $v^{(1)}, \dots, v^{(k)} \in \mathcal{V}$  aufgespannt wird,

$$r(v^{(1)}, \dots, v^{(k)}) := \{[v^{(i)}]_\ell : 0 \leq \ell \leq |v^{(i)}|, 1 \leq i \leq k\}.$$

Ein Knoten  $w$  aus diesem aufgespannten Skelett heißt *essentieller* Knoten falls er der jüngste gemeinsame Vorfahr einer Teilmenge der Knoten  $v^{(1)}, \dots, v^{(k)}$  ist (und somit sind die aufspannenden Knoten  $v^{(i)}$  auch essentielle Knoten). Die Menge der essentiellen Knoten bezeichnen wir durch

$$e(v^{(1)}, \dots, v^{(k)}) := \{v^{(i_1)} \wedge \dots \wedge v^{(i_k)} : i_j \in \{1, \dots, k\}, 1 \leq j \leq k\}.$$

Offensichtlich ist jedes Blatt des Skeletts  $r(v^{(1)}, \dots, v^{(k)})$  einer der aufspannenden Knoten  $v^{(i)}$  sowie jeder Knoten  $v$  in  $r(v^{(1)}, \dots, v^{(k)})$  mit  $c(v, r) \geq 2$  ein essentieller Knoten. Im Allgemeinen lassen sich aber weder die Knoten  $v^{(i)}$  noch die Menge der essentiellen Knoten aus dem Skelett  $r(v^{(1)}, \dots, v^{(k)})$  wiedergewinnen. Uns interessiert die “Gestalt” des Skeletts und die “relative Position” der darin enthaltenen  $k$  ausgezeichneten Knoten  $v^{(1)}, \dots, v^{(k)}$ , d.h. wir notieren nur die Verwandtschaftsverhältnisse der ausgezeichneten Knoten und vergessen deren Altersunterschied. Dazu definieren wir nun eine Äquivalenzrelation auf der Menge der endlichen Teilmengen von  $\mathcal{V}$ : Wir starten zunächst die Menge  $\mathcal{V}$  mit der *lexikographischen Ordnung* aus, d.h. zu  $v \neq w \in \mathcal{V}$  definieren wir

$$v \prec w :\iff v \wedge w = v \quad \text{oder} \quad [v]_{|v \wedge w|+1} < [w]_{|v \wedge w|+1}.$$

Zwei  $k$ -elementige Knotenmengen  $\{v^{(1)}, \dots, v^{(k)}\}$  und  $\{w^{(1)}, \dots, w^{(k)}\}$  heißen *äquivalent*, falls es eine ordnungserhaltende Abbildung  $\varphi$  zwischen den essentiellen Knoten der aufgespannten Skelette gibt, die die Kinderzahlen erhält und

die Menge der ausgezeichneten Knoten ineinander abbildet, d.h. falls eine strikt wachsende Funktion existiert mit

$$\varphi : e(v^{(1)}, \dots, v^{(k)}) \rightarrow e(w^{(1)}, \dots, w^{(k)})$$

so, dass

$$\varphi(v^{(i)}) \in \{w^{(1)}, \dots, w^{(k)}\}, \quad 1 \leq i \leq k,$$

und

$$c(v, r(v^{(1)}, \dots, v^{(k)})) = c(\varphi(v), r(w^{(1)}, \dots, w^{(k)})) \quad (1.10)$$

für jeden Knoten  $v \in e(v^{(1)}, \dots, v^{(k)})$ .

Den kanonischen Repräsentanten  $\{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}$  einer Äquivalenzklasse definieren wir durch die Teilmenge von  $\mathcal{V}$ , die einen Baum minimaler Größe aufspannt. Anschaulich erhält man den kanonischen Repräsentanten, indem man ausgehend von einem beliebigen Skelett der Äquivalenzklasse alle nicht-essentiellen Knoten, die genau ein Kind besitzen, samt Kante entfernt und danach die Knoten neu beschriftet. Im kanonischen Repräsentanten sind nun alle Knoten auch essentielle Knoten. Wir nennen das Tupel

$$(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) := (r(\bar{v}^{(1)}, \dots, \bar{v}^{(k)}), \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \quad (1.11)$$

ein *reduziertes  $k$ -Skelett*. Mit  $\mathcal{R}_k$  bezeichnen wir die Menge aller reduzierten  $k$ -Skelette und  $\mathcal{R}_k^* \subset \mathcal{R}_k$  sei die Menge aller reduzierten  $k$ -Skelette mit  $k$  Blättern. Zur Vereinfachung der Schreibweise werden wir im Folgenden für ein reduziertes  $k$ -Skelett mit  $k$  Blättern auch  $\bar{r}$  anstelle  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})$  schreiben, da man in diesem Fall die reduzierten  $k$ -Skelette mit dem kanonischen Repräsentanten minimaler Größe identifizieren kann. Für alle  $\bar{r} \in \mathcal{R}_k^*$  gilt  $\beta_1(\bar{r}) = 0$ , das heißt  $\mathcal{R}_k^*$  ist die Menge aller verwurzelten, geordneten Bäume mit genau  $k$  Blättern und inneren Knoten die mindestens 2 Kindern besitzen.

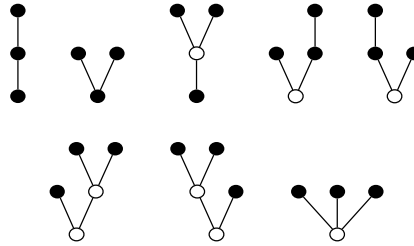


ABBILDUNG 1.3: Menge aller reduzierten 3-Skelette aus  $\mathcal{R}_3$ . In der unteren Reihe sind alle Skelette aus  $\mathcal{R}_3^*$  aufgeführt.

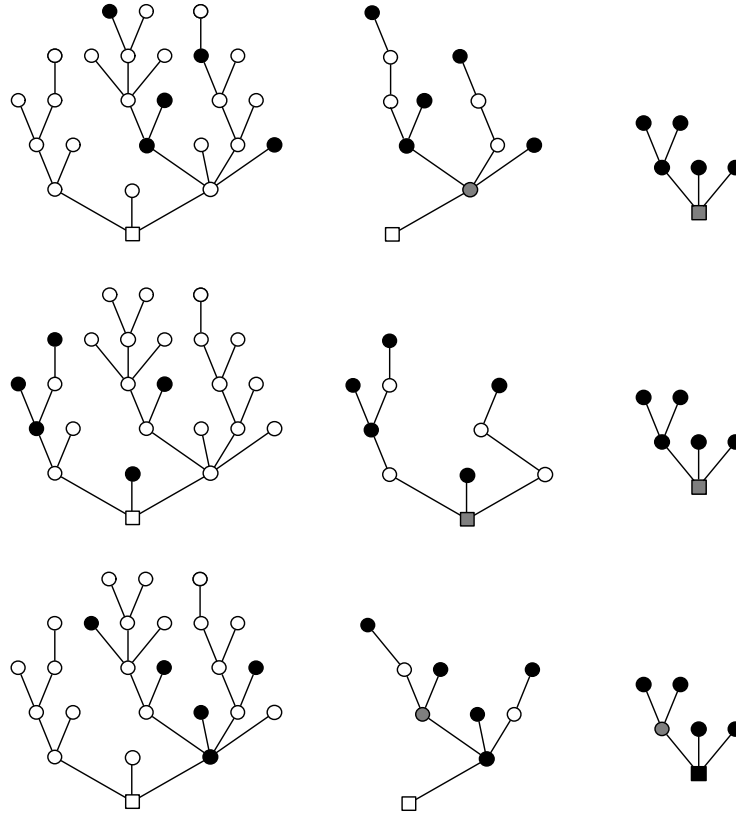


ABBILDUNG 1.2: Beispiele eines Baumes mit  $k = 5$  ausgezeichneten Knoten. Die ausgefüllten Kreise stellen dabei die ausgezeichneten Knoten und die Quadrate die Wurzel dar. Daneben die entsprechenden, durch die markierten Knoten aufgespannten, Skelette. Die essentiellen Knoten sind darin grau schattiert. In der letzte Spalte sind die dazugehörigen reduzierten  $k$ -Skelette aufgeführt. Diese sind alle in  $\mathcal{R}_5 \setminus \mathcal{R}_5^*$  enthalten. Man sieht, dass die reduzierten  $k$ -Skelette der ersten beiden Beispiele übereinstimmen.





## Kapitel 2

# Größenverzerrte Galton-Watson Bäume

In den ersten vier Abschnitten untersuchen wir die Struktur “großer” GW-Bäume indem wir eine Familie von Verteilungen  $\mathcal{L}_{\theta,k}$  von zufälligen Stammbäumen betrachten. Diese Verteilungen stimmen, wenn man sie auf Gesamtgröße bedingt, mit der auf Gesamtgröße bedingten Verteilung eines GW-Baumes überein. Zunächst definieren wir formal die Verteilung  $\mathcal{L}_{\theta,k}(\hat{T})$  durch einen Maßwechsel aus der Verteilung  $\mathcal{L}(T)$  des unverzerrten Baumes. Danach geben wir eine mehrstufige Konstruktion von  $\hat{T}$  entlang eines zufälligen Skeletts an. Mit Hilfe dieser Konstruktion können wir bekannte Resultate über kritische GW-Bäume endlicher Varianz (Verteilung zufällig verwurzelte Teilbäume, Höhe eines uniform gezogenen Knoten, Gesamtkantenlänge, Struktur des bedingten Baumes) in neuem Licht betrachten, sowie neue Ergebnisse im Fall unendlicher Varianz herleiten. Als analytisches Hilfsmittel wird das asymptotische Verhalten der faktoriellen Momente der Kinderzahlverteilung im Falle endlicher sowie unendlicher Varianz zur Verfügung gestellt.

In Abschnitt 2.5 untersuchen wir im Licht des Größenverzerrrens eine Familie von Verteilungen  $\mathcal{L}_{k,n}(\hat{T})$  von zufälligen Stammbäumen, die wir durch eine Größenverzerrung in der  $n$ -ten Generation aus der Verteilung  $\mathcal{L}(T)$  gewinnen. Dieser Ansatz liefert Erkenntnisse über die Höhe des MRCA von  $k$  rein zufällig gezogener Knoten  $n$ -ter Generation, sowie die Struktur des von den  $k$  Knoten aufgespannten Skeletts.

## 2.1 Größenverzerrung in der Gesamtgröße

Die Kinderzahlen der Knoten des klassischen GW-Verzweigungsprozesses sind unabhängige, identisch verteilte Zufallsgrößen. Für jeden endlichen Baum  $t$  gilt daher

$$\mathbb{P}(T = t) = \prod_{v \in t} p_{c(v,t)} = \prod_{j \geq 0} p_j^{\beta_j(t)}. \quad (2.1)$$

Nach Theorem (1.1) ist die Verteilung von  $T$  durch (2.1) für  $\mu \leq 1$  und  $p_1 < 1$  eindeutig bestimmt. Im weiteren Verlauf betrachten wir eine kritische, nichtdegenerierte Kinderzahlverteilung, d.h.  $\mu = 1$  und  $p_1 < 1$ . Diese Voraussetzung wird aber keine wesentliche Einschränkung sein (siehe Bemerkung nach Formel (2.5)). Wir definieren nun die Verteilung  $\mathcal{L}_{\theta,k}$  eines zufälligen Stammbaumes  $\hat{T}$  und beschreiben zunächst den Fall  $k = 0$ . Die Verteilung des Baumes  $\hat{T}$  unter  $\mathcal{L}_{\theta,0}$ ,  $0 < \theta \leq 1$  sei durch die Verteilung eines Galton-Watson Baumes mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  gegeben, d.h. wir definieren für jeden endlichen Baum  $t$

$$\mathbb{P}_{\theta,0}(\hat{T} = t) := \prod_{v \in t} p_{c(v,t)}(\theta)$$

mit der “gekippten” (“tilted”) Kinderzahlverteilung

$$p_j(\theta) := \frac{\theta^j p_j}{f(\theta)}, \quad j \geq 0. \quad (2.2)$$

Dabei bezeichnet  $f$  die Erzeugendenfunktion der Verteilung  $(p_j)_{j \geq 0}$ ,

$$f(s) := \sum_{j=0}^{\infty} p_j s^j, \quad 0 \leq s \leq 1.$$

Die Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  besitzt die Erzeugendenfunktion

$$f_{\theta}(s) := \frac{f(\theta s)}{f(\theta)}, \quad 0 \leq s \leq 1 \quad (2.3)$$

und den Erwartungswert

$$\mu(\theta) := f'_{\theta}(1) = \frac{\theta f'(\theta)}{f(\theta)}. \quad (2.4)$$

Aus (2.3) erkennt man, dass  $(p_j(\theta))_{j \geq 0}$  endliche Momente beliebiger Ordnung besitzt. Weiterhin folgt unmittelbar aus (2.2), dass die Verteilung  $(p_j(\theta))_{j \geq 0}$  in

$\theta$  stochastisch wachsend ist. Speziell ist nun  $\hat{T}$  unter  $\mathcal{L}_{\theta,0}$  für alle  $0 < \theta < 1$  ein subkritischer GW-Baum. Mit Hilfe von (1.7) und (2.1) gilt

$$\mathbb{P}_{\theta,0}(\hat{T} = t) = \prod_{v \in t} \frac{\theta^{c(v,t)} p_{c(v,t)}}{f(\theta)} = \frac{\theta^{s(t)-1}}{f(\theta)^{s(t)}} \mathbb{P}(T = t), \quad 0 < \theta \leq 1, \quad (2.5)$$

d.h.  $\mathcal{L}_{\theta,0}(\hat{T})$  hat eine Dichte bezüglich  $\mathcal{L}(T)$ , welche nur von der Größe des Baumes abhängt. Dies zeigt, dass die Verteilungen der auf Gesamtgröße bedingten GW-Bäume, für Kinderzahlverteilungen, die aus derselben exponentiellen Familie stammen, übereinstimmen (siehe [22]). Daher ist die im weiteren Verlauf vorausgesetzte Kritizität ( $\mu = 1$ ) der Kinderzahlverteilung  $(p_j)_{j \geq 0}$  keine wesentliche Einschränkung.

Wir definieren nun die Verteilung  $\mathcal{L}_{\theta,k}$ , für beliebiges  $k \in \mathbb{N}$ , als *Größenverzerrung* der Verteilung  $\mathcal{L}_{\theta,0}$  über die Gesamtgröße des Baumes

$$\mathbb{P}_{\theta,k}(\hat{T} = t) := \frac{\binom{s(t)}{k} \mathbb{P}_{\theta,0}(\hat{T} = t)}{\mathbb{E}_{\theta,0} \left( \binom{s(\hat{T})}{k} \right)} = \frac{[s(t)]_k \mathbb{P}_{\theta,0}(\hat{T} = t)}{\mathbb{E}_{\theta,0} [s(\hat{T})]_k}, \quad (2.6)$$

mit  $[n]_k := n(n-1) \cdots (n-k+1)$  für  $n, k \in \mathbb{N}$  und  $[n]_0 := 1$ .

Offensichtlich erhält man mit (2.6) ein Wahrscheinlichkeitsmaß auf der Menge der endlichen Stammbäume, das mehr Masse auf große Bäume legt: ein Baum der Größe  $n$  wird unter  $\mathcal{L}_{\theta,k}$  mit dem Faktor  $\bar{c}_{\theta,k}[n]_k$  häufiger gezogen als unter dem GW-Maß  $\mathcal{L}_{\theta,0}$  (mit der Normierung  $\bar{c}_{\theta,k}^{-1} = \mathbb{E}_{\theta,0}[s(\hat{T})]_k$ ). Speziell hat der Baum  $\hat{T}$  unter  $\mathcal{L}_{\theta,k}$  nun eine Mindestgröße von  $k$  Knoten und die Kinderzahlen sind für  $k \geq 1$  nicht mehr unabhängig. Aus (2.5) und (2.6) sieht man, dass auch  $\mathcal{L}_{\theta,k}(\hat{T})$  eine Dichte bezüglich  $\mathcal{L}(T)$  besitzt, die nur von der Größe des Baumes abhängt,

$$\mathbb{P}_{\theta,k}(\hat{T} = t) = \frac{[s(t)]_k}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} \frac{\theta^{s(t)-1}}{f(\theta)^{s(t)}} \mathbb{P}(T = t). \quad (2.7)$$

Daraus folgt unmittelbar Eigenschaft (1.2):

Die auf Gesamtgröße bedingte Verteilung ist unabhängig von den Parametern  $\theta$  und  $k$ , denn

$$\mathcal{L}_{\theta,k}(\hat{T} \mid s(\hat{T}) = n) = \mathcal{L}(T \mid s(T) = n), \quad n \geq 1.$$

Zum Beweis der Monotonie der Funktion  $h$  (Eigenschaft (1.3)), stellen wir zunächst mittels (2.7) fest, dass

$$\mathbb{P}_{\theta,k}(s(\hat{T}) = n) = \begin{cases} \frac{[n]_k}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} \frac{\theta^{n-1}}{f(\theta)^n} \mathbb{P}(s(T) = n), & \text{falls } n \geq k, \\ 0, & \text{sonst.} \end{cases} \quad (2.8)$$

Daraus folgt

$$\frac{\mathbb{P}_{\theta,k}(s(\hat{T}) = n+1)}{\mathbb{P}_{\theta,k}(s(\hat{T}) = n)} = \frac{n+1}{n-k+1} \frac{\theta}{f(\theta)} \frac{\mathbb{P}(s(T) = n+1)}{\mathbb{P}(s(T) = n)}, \quad n \geq k.$$

Für jedes  $n$  wächst der Faktor  $(n+1)/(n-k+1)$  in  $k$  und für jedes  $k$  der Faktor  $\theta/f(\theta)$  in  $\theta$  (da  $f$  konvex und  $f'(1) = 1$ ). Dies zeigt, dass die Verteilung  $\mathcal{L}_{\theta,k}(s(\hat{T}))$  stochastisch von  $\mathcal{L}_{\theta',k'}(s(\hat{T}))$  dominiert wird, falls  $\theta \leq \theta'$  und  $k \leq k'$ . Im speziellen folgt damit die behauptete Monotonie von  $h$ . Weiterhin ist die Folge der reskalierten Generationsgrößen  $(\mu(\theta)^{-m} |z_m(\hat{T})|)_{m \geq 0}$  unter  $\mathcal{L}_{\theta,0}$  ein Martingal und es gilt

$$h(\theta, 0) = \mathbb{E}_{\theta,0} s(\hat{T}) = \sum_{m=0}^{\infty} \mathbb{E}_{\theta,0} |z_m(\hat{T})| = \frac{1}{1 - \mu(\theta)}, \quad 0 < \theta < 1. \quad (2.9)$$

Eigenschaft (1.4) folgt nun aus (2.4), (2.9) und der Monotonie von  $h$  in  $k$ . In Abschnitt 2.4 werden wir das asymptotische Verhalten von  $h$  ausführlicher untersuchen.

## 2.2 Konstruktion des größenverzerrten Baumes $\hat{T}$ entlang eines zufälligen Skeletts

Der größenverzerrte Baum  $\hat{T}$  besitzt unter  $\mathcal{L}_{\theta,k}$  nicht die einfache Unabhängigkeitsstruktur eines GW-Baumes - daher ist eine Konstruktion der Zufallsvariable  $\hat{T}$  unter  $\mathcal{L}_{\theta,k}$  auf den ersten Blick nicht direkt möglich. Einen Ausweg eröffnet die Betrachtung des Paares  $(\hat{T}, B_k)$ , wobei  $B_k = \{V^{(1)}, \dots, V^{(k)}\}$  eine Menge von  $k$  verschiedenen, rein zufällig aus der Knotenmenge von  $\hat{T}$  gezogenen Knoten ist (d.h. gegeben  $\{\hat{T} = t\}$  ist  $B_k$  gleichverteilt auf allen  $\binom{s(t)}{k}$   $k$ -elementigen Teilmengen der Knoten von  $t$ ). Die Verteilung von  $\hat{T}$  erhält man damit als Randverteilung dieser multivariaten Zufallsvariable auf dem Raum der Bäume. Die folgenden Beobachtungen führen zu einer einfachen und anschaulichen mehrstufigen Konstruktion des Baumes  $\hat{T}$  entlang eines zufälligen,

durch  $B_k$  aufgespannten Skeletts. Sei nun  $k \geq 1$  und  $0 < \theta < 1$ . Die Verteilung von  $(\hat{T}, B_k)$  unter  $\mathcal{L}_{\theta,k}$  ist für einen endlichen Baum  $t$  der Größe  $s(t) \geq k$  und einer beliebigen  $k$ -elementigen Knotenmenge  $\{v^{(1)}, \dots, v^{(k)}\} \subset t$  gegeben durch

$$\mathbb{P}_{\theta,k}(\hat{T} = t, B_k = \{v^{(1)}, \dots, v^{(k)}\}) = \binom{s(t)}{k}^{-1} \mathbb{P}_{\theta,k}(\hat{T} = t). \quad (2.10)$$

Mit (2.6) erhalten wir

$$\mathbb{P}_{\theta,k}(\hat{T} = t, B_k = \{v^{(1)}, \dots, v^{(k)}\}) = \frac{k!}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} \mathbb{P}_{\theta,0}(\hat{T} = t). \quad (2.11)$$

Zum besseren Verständnis beschreiben wir die Konstruktion zunächst für  $k = 1$ . In diesem Fall besteht das Skelett aus einer einzigen Ahnenlinie und der kanonische Repräsentant ist der Baum, der nur die Wurzel enthält.

### Verteilung des Tupels $(\hat{T}, B_1)$

Sei  $t$  ein endlicher Baum und  $v$  ein Knoten aus  $t$ . Wir stellen nun (2.11) folgendermaßen dar (vgl. (2.9))

$$\begin{aligned} \mathbb{P}_{\theta,1}(\hat{T} = t, B_1 = \{v\}) &= (1 - \mu(\theta)) \prod_{w \in t} p_{c(w,t)}(\theta) \\ &= (1 - \mu(\theta)) \mu(\theta)^{|v|} \prod_{\ell=0}^{|v|-1} \frac{c([v]_{\ell}, t) p_{c([v]_{\ell}, t)}(\theta)}{\mu(\theta)} \frac{1}{c([v]_{\ell}, t)} \\ &\quad \prod_{w \notin \{[v]_{\ell}: 0 \leq \ell \leq |v|-1\}} p_{c(w,t)}(\theta). \end{aligned}$$

Aus dieser Darstellung erkennt man, dass die Höhe des ausgezeichneten Knotens geometrisch verteilt ist mit Mittelwert  $\mu(\theta)/(1 - \mu(\theta))$ . Weiterhin gehören der ausgezeichnete Knoten und seine Vorgänger (bis auf die Wurzel) unabhängigen größenverzerrten Familien an, d.h. die Wahrscheinlichkeit, dass solch ein Knoten sich unter  $j$  Geschwistern befindet, ist  $j p_j(\theta)/\mu(\theta)$ ,  $j \geq 1$ . Die relativen Positionen der Knoten dieser ausgezeichneten Ahnenlinie sind bedingt gleichverteilt unter den Geschwistern und unabhängig von der Generation. Schließlich erkennt man am letzten Ausdruck der Darstellung, dass die Teilbäume, verwurzelt in den Knoten außerhalb der Ahnenlinie, unabhängige Kopien von GW-Bäumen mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  sind.

**Bemerkungen:** Die Struktur des Baumes  $\hat{T}$  mit einem uniform gezogenen Knoten  $V$  lässt sich intuitiv leicht erklären: Ziehe aus einem zufälligen “Wald” von  $N$  (mit  $N \rightarrow \infty$ ) unabhängigen GW-Bäumen mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  uniform unter allen Knoten des Waldes einen Knoten  $V$  und betrachte den damit ausgewählten Baum. Offenbar haben große Bäume eine größere Chance gezogen zu werden (dies ist ein allgemein bekanntes Phänomen der Größenverzerrung wie es z.B. beim Wartezeitenparadoxon (siehe z.B. [13]) auftritt). Genauer gilt: Der gezogene Baum besitzt die größenverzerrte Verteilung  $\mathcal{L}_{\theta,1}$  und der ausgezeichnete Knoten  $V$  ist uniform auf den Knoten des gezogenen Baumes verteilt. Da im Wald die Anzahl der Familien der Größe  $j$  mit einer Häufigkeit vorkommen, die asymptotisch proportional zu  $j p_j(\theta)$  ist, sind die Größen der Familien in der ausgezeichneten Ahnenlinie größenverzerrt. Nach dem Gesetz der großen Zahlen gibt es im Wald approximativ  $N \mathbb{P}_{\theta,0}(\hat{T} = t)$  viele endliche Bäume  $t$ . Darunter befinden sich nach (2.9) approximativ  $N(1 - \mu(\theta))^{-1} \mathbb{P}_{\theta,0}(\hat{T} = t')$  viele Teilbäume  $t'$ . Somit ist die Wahrscheinlichkeit, dass ein zufällig gezogener Knoten eine Wurzel ist, approximativ  $1 - \mu(\theta)$ , unabhängig des daraus wachsenden Teilbaumes.

Die Zerlegung von  $\hat{T}$  entlang der Ahnenlinie eines zufälligen Knotens im Spezialfall der Poissonverteilung wurde in [4] hergeleitet. Die analoge Konstruktion in kontinuierlicher Zeit für binäre Verzweigungsprozesse findet man in [14].

### Verteilung des Tupels $(\hat{T}, B_k)$

Wir beschreiben nun die Verteilung im Fall  $k \geq 1$ . Dazu ist es nützlich, folgende Notation einzuführen: Sei  $(p_j(\theta, i))_{j \geq 0}$  für  $i \in \mathbb{N}_0$  die, mit der Funktion  $[\cdot]_i$  aus  $(p_j(\theta))_{j \geq 0}$  erhaltene, größenverzerrte Verteilung

$$p_j(\theta, i) := \begin{cases} \frac{[j]_i p_j(\theta)}{\mu(\theta, i)}, & \text{für } j \geq i, \\ 0, & \text{sonst.} \end{cases} \quad (2.12)$$

Die Normierung  $\mu(\theta, i)$  ist dabei das  $i$ -te faktorielle Moment von  $(p_j(\theta))_{j \geq 0}$ . Mit (2.2) lässt sich dies durch die Erzeugendenfunktion  $f$  darstellen

$$\mu(\theta, i) := \sum_{j=i}^{\infty} [j]_i p_j(\theta) = \frac{\theta^i f^{(i)}(\theta)}{f(\theta)}. \quad (2.13)$$

Dabei ist  $p_j(\theta, 0) = p_j(\theta)$  und  $\mu(\theta, 0) = 1$  für jedes  $\theta$ . Wir nennen  $(p_j(\theta, i))_{j \geq 0}$  die  $i$ -fach größenverzerrte Verteilung von  $(p_j(\theta))_{j \geq 0}$ . Wie im Fall  $k = 1$  stellen wir Formel (2.11) um zu

$$\begin{aligned}
\mathbb{P}_{\theta, k}(\hat{T} = t, B_k = \{v^{(1)}, \dots, v^{(k)}\}) &= \frac{k!}{\mathbb{E}_{\theta, 0}[s(\hat{T})]_k} \prod_{v \in t} p_{c(v, t)}(\theta) \\
&= \frac{k!}{(1 - \mu(\theta))^{|e|} \mathbb{E}_{\theta, 0}[s(\hat{T})]_k} \prod_{v \in e} \frac{\mu(\theta, c(v, r))}{c(v, r)!} \\
&\quad \cdot (1 - \mu(\theta))^{|e|} \prod_{v \in r \setminus e} \frac{\mu(\theta, c(v, r))}{c(v, r)!} \\
&\quad \cdot \prod_{v \in r} \frac{[c(v, t)]_{c(v, r)} p_{c(v, t)}(\theta)}{\mu(\theta, c(v, r))} \left( \frac{c(v, t)}{c(v, r)} \right)^{-1} \prod_{v \in t \setminus r} p_{c(v, t)}(\theta) \\
&= \frac{k!}{(1 - \mu(\theta))^{|e|} \mathbb{E}_{\theta, 0}[s(\hat{T})]_k} \prod_{v \in e} \frac{\mu(\theta, c(v, r))}{c(v, r)!} \\
&\quad \cdot (1 - \mu(\theta))^{|e|} \mu(\theta)^{|r \setminus e|} \\
&\quad \cdot \prod_{v \in r} p_{c(v, t)}(\theta, c(v, r)) \left( \frac{c(v, t)}{c(v, r)} \right)^{-1} \prod_{v \in t \setminus r} p_{c(v, t)}(\theta), \quad (2.14)
\end{aligned}$$

mit  $r = r(v^{(1)}, \dots, v^{(k)})$  und  $e = e(v^{(1)}, \dots, v^{(k)})$ .

Bei der letzten Gleichheit wurde die Eigenschaft benutzt, dass jeder nicht-essentielle Knoten  $v \in r \setminus e$  nur ein Kind in  $r$  besitzt. Nach Definition (1.10) erkennt man, dass der Ausdruck

$$(1 - \mu(\theta))^{-|e|} \prod_{v \in e} \frac{\mu(\theta, c(v, r))}{c(v, r)!}$$

in (2.14) nur von der Form des von den Knoten  $v^{(1)}, \dots, v^{(k)}$  aufgespannten Skeletts abhängt und für äquivalente  $k$ -Skelette denselben Wert liefert. Mit Hilfe des kanonischen Repräsentanten  $\bar{r} = r(\bar{v}^{(1)}, \dots, \bar{v}^{(k)})$  lässt sich dieser Ausdruck auch in folgender Form, auf die wir im weiteren Verlauf zurückgreifen werden, schreiben

$$(1 - \mu(\theta))^{-s(\bar{r})} \prod_{i \geq 1} \left( \frac{\mu(\theta, i)}{i!} \right)^{\beta_i(\bar{r})}. \quad (2.15)$$

## Konstruktion des größenverzerrten Baumes

Identität (2.14) motiviert zur folgenden mehrstufigen Konstruktion einer multivariaten Zufallsvariable mit Verteilung (2.10):

**Konstruktion 2.1.** Sei  $k \geq 1$  und  $0 < \theta < 1$ .

1. **Reduziertes  $k$ -Skelett.** Ziehe ein zufälliges reduziertes  $k$ -Skelett  $\bar{R}_k$  mit Verteilung

$$\begin{aligned} \mathbb{P}_\theta(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) \\ = c_{\theta,k} (1 - \mu(\theta))^{-s(\bar{r})} \prod_{i \geq 1} \left( \frac{\mu(\theta, i)}{i!} \right)^{\beta_i(\bar{r})} \end{aligned} \quad (2.16)$$

für  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k$ , mit  $c_{\theta,k} = k!/\mathbb{E}_{\theta,0}[s(\hat{T})]_k$ .

2.  **$k$ -Skelett.** Seien  $L_v, v \in \mathcal{V}$  unabhängige, geometrisch zum Parameter  $1 - \mu(\theta)$  verteilte Zufallsvariablen. Gegeben  $\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})$  füge zu jedem Knoten  $v \in \bar{r} \setminus \{\Delta\}$  eine Kette von  $L_v$  Knoten zwischen  $v$  und seinem Vorgänger ein. Bringe zusätzlich eine Kette von  $L_\Delta$  Knoten an die Wurzel von  $\bar{r}$  an. Das  $k$ -Skelett  $R_k$  mit  $k$  ausgezeichneten Knoten erhält man nun durch Neubeschriftung der Knoten.
3. **Aufgedicktes  $k$ -Skelett.** Gegeben  $R_k = r$ , ziehe für jeden Knoten  $v \in r \setminus \Lambda(r)$  Kinderzahlen  $X_v$  gemäß einer  $c(v, r)$ -fach größenverzerrten Kinderzahlverteilung und gebe den  $X_v - c(v, r)$  neuen Kinder rein zufällig eine der  $\binom{X_v}{c(v, r)}$  möglichen Anordnungen.
4. **Aufforsten des aufgedickten Skeletts.** Füge an jedes Blatt des aufgedickten Skeletts einen GW-Baum mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  an.

Wir notieren das durch die Konstruktion gewonnene Tupel, bestehend aus dem zufälligen Baum und der darin enthaltenen Menge von  $k$  ausgezeichneten Knoten, durch  $(T_k^*, B_k^*)$ . Die folgende Proposition ist nun eine unmittelbare Folgerung aus der Darstellung von  $\mathcal{L}_{\theta,k}(\hat{T}, B_k)$  in (2.14).

**Proposition 2.2.** Für jedes  $k \geq 1$  und  $0 < \theta < 1$  gilt

$$\mathcal{L}_\theta(T_k^*, B_k^*) = \mathcal{L}_{\theta,k}(\hat{T}, B_k).$$



## Bemerkungen und Anwendungen

**Zufällig verwurzelter Teilbaum (“Fringe Subtree”).** Aus obiger Zerlegung von  $\hat{T}$  folgt unmittelbar, dass der an einem zufällig gezogenen Knoten verwurzelte Teilbaum ein GW-Baum mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  ist. Dies erklärt ein Resultat in [1] (Lemma 9), welches zeigt, dass der rein zufällig verwurzelte Teilbaum, in einem auf Gesamtgröße  $n$  bedingten, kritischen GW-Baum endlicher Varianz, asymptotisch für  $n \rightarrow \infty$  die Verteilung  $\mathcal{L}(T)$  besitzt.

**Höhe eines uniform gezogenen Knotens.** Die reskalierte Höhe  $|V_1|$  ist unter  $\mathcal{L}_{\theta,1}$  im Limes  $\theta \uparrow 1$  exponentialverteilt. Dieser Grenzwert unterscheidet sich von der asymptotischen Verteilung der Höhe eines uniform gezogenen Knotens im kritischen, auf Gesamtgröße  $n$  bedingten GW-Baumes endlicher Varianz. Diese hat die Dichte (Theorem 4.5 in [26])

$$g(x) = x \exp(-x^2/2), \quad x \geq 0. \quad (2.17)$$

Wir werden in Abschnitt 2.4 sehen, in welchem Verhältnis die Grenzverteilungen zueinander stehen.

**Momente.** Mit Hilfe obiger Konstruktion ist es möglich die faktoriellen Momente der Größe eines GW-Baumes  $\hat{T}$  mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  zu berechnen. Nach (2.6) erhält man

$$\mathbb{E}_{\theta,k} s(\hat{T}) = \sum_t \frac{s(t)[s(t)]_k \mathbb{P}_{\theta,0}(\hat{T} = t)}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} = \frac{\mathbb{E}_{\theta,0}[s(\hat{T})]_{k+1}}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} + k.$$

Dies liefert die rekursive Darstellung

$$\mathbb{E}_{\theta,0}[s(\hat{T})]_{k+1} = (\mathbb{E}_{\theta,k} s(\hat{T}) - k) \mathbb{E}_{\theta,0}[s(\hat{T})]_k, \quad k \geq 0. \quad (2.18)$$

Dabei kann der erste Erwartungswert in (2.18) mit Hilfe der Konstruktion 2.1 berechnet werden. Die Verteilung  $\mathcal{L}_{\theta,k}(s(\hat{T}))$  und deren asymptotisches Verhalten werden wir in Abschnitt 2.4 genauer analysieren.

**Gesamtkantenlänge.** Die Gesamtkantenlänge (Total Path Length)  $d^*(t) := \sum_{v \in t} |v|$  eines endlichen Baumes  $t$  ist die Summe der Generationshöhen aller Knoten. Diese Größe spielt z.B. bei der Analyse von Zugriffsoperationen auf Daten, die in einer Baumstruktur abgespeichert werden, eine wichtige Rolle. Eine zu ihr eng verwandte Größe ist der erwartete Abstand zur Wurzel  $\bar{d}(t) := d^*(t)/s(t)$  eines rein zufällig aus  $t$  gezogenen Knotens. Es gilt im Kontext des Baumes  $\hat{T}$

$$\mathbb{E}_{\theta,0} d^*(\hat{T}) = \sum_t s(t) \bar{d}(t) \mathbb{P}_{\theta,0}(\hat{T} = t) = \mathbb{E}_{\theta,0} s(\hat{T}) \mathbb{E}_{\theta,1} \bar{d}(\hat{T}).$$

Für die zweite Gleichheit wurde Beziehung (2.6) herangezogen. Da nach Konstruktion 2.1 die Höhe eines rein zufällig gezogenen Knotens aus  $\hat{T}$  unter  $\mathcal{L}_{\theta,1}$  geometrisch zum Erwartungswert  $\mu(\theta)/(1-\mu(\theta))$  verteilt ist, erhält man mit (2.9)

$$\mathbb{E}_{\theta,0}d^*(\hat{T}) = \frac{\mu(\theta)}{(1-\mu(\theta))^2}. \quad (2.19)$$

Im Vergleich dazu besitzt die Gesamtkantenlänge des kritischen, auf Gesamtgröße  $n$  bedingten GW-Baumes endlicher Varianz eine Grenzverteilung bei Normierung mit  $n^{\frac{3}{2}}$  (siehe [2, 31]). Obwohl die bedingte Verteilung des größenverzerrten Baumes mit der des unverzerrten Baumes übereinstimmt (siehe (1.2)), wächst die erwartete Gesamtkantenlänge von  $\hat{T}$  unter  $\mathcal{L}_{\theta,0}$  quadratisch (anstelle der Potenz  $\frac{3}{2}$ ) in der erwarteten Größe  $h(\theta,0) = (1-\mu(\theta))^{-1}$  des Baumes. Dies zeigt, dass die besonders großen Bäume den Erwartungswert in (2.19) in hohem Maße bestimmen.

### 2.3 Limes des reduzierten Baumes

Wir beschreiben nun die asymptotische Struktur des Skeletts, aufgespannt von  $k$  rein zufällig und ohne Zurücklegen aus  $\hat{T}$  gezogenen Knoten.

#### Endliche Varianz der Kinderzahlverteilung

Zuerst betrachten wir eine kritische Verteilung  $(p_j)_{j \geq 0}$  mit endlicher Varianz

$$\sigma^2 := \sum_{j \geq 0} j(j-1)p_j = \lim_{\theta \uparrow 1} f^{(2)}(\theta).$$

Es stellt sich heraus, dass im Limes  $\theta \uparrow 1$  das von  $k$  rein zufällig und ohne Zurücklegen aus  $\hat{T}$  gezogenen Knoten aufgespannte und anschließend reduzierte Skelett (alle Knoten des Skeletts mit nur einem Nachfolger werden samt ausgehender Kante entfernt) ein vollständig binärer Baum mit  $k$  Blättern ist. Mit Proposition 2.2 lässt sich diese Aussage mittels des reduzierten  $k$ -Skeletts  $\bar{R}_k$  aus Konstruktion 2.1 formulieren.

**Theorem 2.3.** *Sei  $\mu = 1$  und  $0 < \sigma^2 < \infty$ . Dann gilt für jedes  $k \geq 2$ ,*

$$\begin{aligned} & \lim_{\theta \uparrow 1} \mathbb{P}_{\theta}(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) \\ &= \begin{cases} C_{k-1}^{-1}, & \text{für } (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k^* \text{ und } \beta_2(\bar{r}) = k-1, \\ 0, & \text{sonst,} \end{cases} \end{aligned} \quad (2.20)$$

wobei  $C_n = \frac{1}{2n+1} \binom{2n+1}{n}$  die  $n$ -te Catalan Zahl ist.

**Beweis.** Nach (2.16) gilt für jedes reduzierte  $k$ -Skelett  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k$

$$\mathbb{P}_\theta(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) = c_{\theta,k} (1 - \mu(\theta))^{-s(\bar{r})} \prod_{i \geq 1} \left( \frac{\mu(\theta, i)}{i!} \right)^{\beta_i(\bar{r})} \quad (2.21)$$

Da  $\sigma^2 < \infty$  gilt asymptotisch (siehe (2.13))

$$1 - \mu(\theta) \sim \sigma^2(1 - \theta) \quad \text{und} \quad \mu(\theta, 2) \rightarrow \sigma^2 < \infty \quad \text{für} \quad \theta \uparrow 1. \quad (2.22)$$

Somit erhält man für zwei reduzierte  $k$ -Skelette  $(\bar{r}_1, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})$  und  $(\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\}) \in \mathcal{R}_k$

$$\begin{aligned} & \log \frac{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\}))}{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_1, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}))} \\ &= (s(\bar{r}_1) - s(\bar{r}_2)) \log(1 - \theta) + \sum_{i \geq 3} (\beta_i(\bar{r}_2) - \beta_i(\bar{r}_1)) \log \mu(\theta, i) + O(1). \end{aligned} \quad (2.23)$$

Sei nun  $\bar{r}_1$  ein vollständiger binärer Baum mit  $k$  Blättern, d.h.  $\bar{r}_1 \in \mathcal{R}_k^*$  und  $\beta_i(\bar{r}_1) = 0$  für  $i \geq 3$ . Durch Umformen von (2.23) erhält man

$$\begin{aligned} & \log \frac{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\}))}{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_1, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}))} \\ &= (2k - 1 - s(\bar{r}_2)) \log(1 - \theta) + \sum_{i \geq 3} \beta_i(\bar{r}_2) \log \mu(\theta, i) + O(1) \\ &= \left( 2k - 1 - \left( s(\bar{r}_2) + \sum_{i \geq 3} (i - 2) \beta_i(\bar{r}_2) \right) \right) \log(1 - \theta) \\ &\quad + \sum_{i \geq 3} \beta_i(\bar{r}_2) \log (\mu(\theta, i) (1 - \theta)^{i-2}) + O(1). \end{aligned} \quad (2.24)$$

Wir zeigen nun, dass für jeden nicht vollständig binären Baum  $r_2$  die rechte Seite in (2.24) mit  $\theta \uparrow 1$  gegen  $-\infty$  strebt:

Für den ersten Term in (2.24) stellen wir fest, dass für jedes reduzierte  $k$ -Skelett  $(\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\}) \in \mathcal{R}_k$  gilt  $\beta_0(\bar{r}_2) + \beta_1(\bar{r}_2) \leq k$ . Mit Hilfe von Identität (1.8) erhält man

$$\begin{aligned} s(\bar{r}_2) + \sum_{i \geq 3} (i - 2) \beta_i(\bar{r}_2) &= \beta_0(\bar{r}_2) + \beta_1(\bar{r}_2) + \sum_{i \geq 1} (i - 1) \beta_i(\bar{r}_2) \\ &= 2\beta_0(\bar{r}_2) + \beta_1(\bar{r}_2) - 1 \leq 2k - 1. \end{aligned} \quad (2.25)$$

Dabei gilt in (2.25) die Gleichheit genau dann, wenn  $\beta_0(\bar{r}_2) = k$ .

Das asymptotische Verhalten der zweiten Summe in (2.24) untersuchen wir mittels einer Taylorentwicklung von  $f^{(2)}(s)$  um  $\theta$ . Es gilt

$$0 \leq \frac{f^{(i)}(\theta)}{(i - 2)!} (1 - \theta)^{i-2} \leq f^{(2)}(1) - f^{(2)}(\theta), \quad i \geq 3.$$

Mittels (2.13) folgt

$$\lim_{\theta \uparrow 1} (1 - \theta)^{i-2} \mu(\theta, i) = 0, \quad i \geq 3. \quad (2.26)$$

Für alle reduzierten  $k$ -Skelette  $(\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\})$ , die nicht vollständig binär sind, erhält man somit

$$\lim_{\theta \uparrow 1} \log \frac{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_2, \{\bar{w}^{(1)}, \dots, \bar{w}^{(k)}\}))}{\mathbb{P}_\theta(\bar{R}_k = (\bar{r}_1, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}))} = -\infty.$$

Da die Gewichte in (2.21) nur durch die Kinderzahlen  $(\beta_i(\bar{r}))_{i \geq 0}$  bestimmt werden, haben alle vollständig binären Bäume mit  $k$  Blättern, von denen es  $C_{k-1} = \frac{1}{2k-1} \binom{2k-1}{k}$  gibt, die gleiche Wahrscheinlichkeit.

□

## Unendliche Varianz der Kinderzahlverteilung

Wir untersuchen nun die asymptotische Struktur des reduzierten Skeletts für kritische Verteilung  $(p_j)_{j \geq 0}$  mit unendlicher Varianz. Dazu müssen wir eine gewisse Regularitätsbedingung an die Schwanzwahrscheinlichkeit der Kinderzahlverteilung stellen. Eine natürliche Bedingung, die bei der Analyse von Summen von unabhängigen, identisch verteilten Zufallsvariablen benötigt wird, ist, dass die Verteilung der Zufallsvariablen im Anziehungsbereich einer stabilen Verteilung liegt. Die kritische Kinderzahlverteilung  $(p_j)_{j \geq 0}$  liegt im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung  $G_\alpha$ , falls es Zahlen  $a_n > 0$  gibt, so dass  $a_n^{-1}(X_1 + \dots + X_n - n)$  in Verteilung gegen  $G_\alpha$  konvergiert. Dabei nennt man eine Verteilung  $G$  stabil, wenn es für unabhängige Zufallsvariablen  $X, X_1, X_2$  mit Verteilung  $G$  reelle Zahlen  $a, b$  gibt, so dass  $\mathcal{L}(X_1 + X_2) = \mathcal{L}(aX + b)$ . Man kann zeigen (siehe z.B. [13]), dass  $G_\alpha$  eine Dichte besitzt (zu der man im allgemeinen Fall aber nur Zugang durch eine Reihendarstellung hat) und alle Momente der Ordnung kleiner als  $\alpha$  existieren.

Da wir hier kritische Verteilungen betrachten, ist  $\alpha \in (1, 2]$  und  $a_n = o(n)$ . Weiterhin ist die Grenzverteilung  $G_\alpha$  für  $1 < \alpha < 2$  eine spektral-positive Verteilung. In unserem Kontext bedeutet dies, dass der relative Anteil der Masse der Randverteilungen nur auf der positiven Achse liegt, d.h. für die Verteilungsfunktion  $F_\alpha$  von  $G_\alpha$  gilt  $F_\alpha(-x)/(1 - F_\alpha(x) + F_\alpha(-x)) \rightarrow 0$  für  $x \rightarrow \infty$ . Genauer kann man zeigen, dass  $F_\alpha(x)$  für  $x \rightarrow -\infty$  exponentiell schnell fällt, hingegen

im Positiven  $1 - F_\alpha(x)$  regulär variiert (wobei es sich bei  $G_2$  um die Normalverteilung handelt). Die Grenzverteilung  $G_\alpha$  ist bis auf eine Skalierungskonstante eindeutig.

Eine Verteilung liegt im sogenannten *normalen* Anziehungsbereich einer stabilen Verteilung  $G_\alpha$  falls die Normierungsfolge  $a_n := n^{1/\alpha}$  gewählt werden kann. Beispielsweise liegt eine stabile Verteilung im normalen Anziehungsbereich von sich selbst. Einen Überblick über stabile Verteilungen und deren Anziehungsbereiche findet man in Kapitel XVII.5 in [13] oder 8.3 in [8].

Im weiteren Verlauf sei  $(p_j)_{j \geq 0}$  eine kritische Verteilung, die im Anziehungsbereich einer stabilen Verteilung  $G_\alpha$  mit Index  $\alpha$  liegt.

Wie im Fall endlicher Varianz werden wir im Fall unendlicher Varianz sehen, dass unter  $\mathcal{L}_{\theta,k}$  im Limes  $\theta \uparrow 1$  die  $k$  ausgezeichneten Knoten Blätter sind. Nun wird sich aber herausstellen, dass auch reduzierte Skelette, deren Knoten mehr als 2 Kinder besitzen, im Limes mit positiver Wahrscheinlichkeit bestehen bleiben. Wir formulieren dieses Ergebnis wieder mit Hilfe des reduzierten  $k$ -Skeletts aus Konstruktion 2.1.

**Theorem 2.4.** *Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt für jedes  $k \geq 2$ ,*

$$\begin{aligned} & \lim_{\theta \uparrow 1} \mathbb{P}_\theta(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) \\ &= \begin{cases} (c_k(\alpha))^{-1} \prod_{j \geq 2} q_j(\alpha)^{\beta_j(\bar{r})}, & \text{für } (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k^*, \\ 0, & \text{sonst,} \end{cases} \end{aligned} \quad (2.27)$$

mit

$$q_j(\alpha) = \frac{\alpha - 1}{\Gamma(2 - \alpha)} \frac{\Gamma(j - \alpha)}{\Gamma(j + 1)}, \quad j \geq 2 \quad (2.28)$$

und

$$c_k(\alpha) = \frac{1}{\Gamma(k + 1)} \prod_{i=1}^{k-1} (i\alpha - 1). \quad (2.29)$$

wobei

$$\Gamma(0)/\Gamma(0) := 1 \quad \text{und} \quad \Gamma(j)/\Gamma(0) := 0 \quad \text{für } j \geq 1.$$

**Bemerkungen.**

**Konsistenz im Fall endlicher Varianz.** Im Fall  $\alpha = 2$  ist  $q_2(2) = \frac{1}{2}$  und  $q_j(2) = 0$  für  $j \geq 3$ . Dies zeigt, dass Theorem 2.4 konsistent mit Theorem 2.3 ist.

**GW-Baum bedingt auf Blätterzahl.** Jedes reduzierte  $k$ -Skelett mit  $k$  Blättern kann mit einem minimalen Baum  $\bar{r}$  der entsprechenden Äquivalenzklasse identifiziert werden. Aus Produktdarstellung (2.27) und

$$\sum_{j=2}^{\infty} q_j(\alpha) = 1 - \frac{1}{\alpha} < 1$$

folgt, dass der kanonische Repräsentant in Theorem 2.4 ein auf  $k$  Blätter bedingter GW-Baum mit Kinderzahlverteilung  $(q_j(\alpha))_{j \geq 0}$  ( $q_0(\alpha) := \frac{1}{\alpha}$  und  $q_1(\alpha) := 0$ ) ist.

**Blätterzahlverteilung.** Erweitert man (2.28) mit  $(q_0(\alpha)/q_0(\alpha))^k$  zu

$$(q_0(\alpha)^k c_k(\alpha))^{-1} \prod_{j \geq 0} q_j(\alpha)^{\beta_j(\bar{r})},$$

so sieht man, dass ein GW-Baum mit Kinderzahlverteilung  $(q_j(\alpha))_{j \geq 0}$  mit einer Wahrscheinlichkeit von  $c_k(\alpha)/\alpha^k$  genau  $k$  Blätter besitzt.

**Stabiler Anziehungsbereich.** Die Verteilung  $(q_j(\alpha))_{j \geq 0}$  liegt im normalen Anziehungsbereich einer stabilen Verteilung  $G_\alpha$ . Die Erzeugendenfunktion ist

$$f_\alpha(s) = s + \frac{1}{\alpha}(1-s)^\alpha, \quad 0 \leq s \leq 1$$

und für die Schwanzwahrscheinlichkeiten gilt

$$\sum_{j \geq n} q_j(\alpha) = \frac{(\alpha-1)\Gamma(n-\alpha)}{\alpha\Gamma(2-\alpha)\Gamma(n)} \sim \frac{\alpha}{\alpha\Gamma(2-\alpha)} n^{-\alpha}.$$

Dies ist der Prototyp einer Verteilung, die im Anziehungsbereich einer stabilen Verteilung liegt. Die langsam variierende Funktion  $L$  ist hier konstant gleich  $q_0(\alpha) = 1/\alpha$ .

T. Duquesne und J.-F. Le Gall erhalten Verteilung (2.27) als Randverteilung eines zeitkontinuierlichen stabilen Verzweigungsprozesses (Theorem 3.3.3 in [10], Theorem 3.2 in [11]).

Zum Beweis von Theorem 2.4 benötigen wir das asymptotische Verhalten der faktoriellen Momente (2.13) für  $\theta \uparrow 1$ . Wir betrachten dazu die Fälle  $1 < \alpha < 2$  und  $\alpha = 2$  getrennt voneinander.

**Lemma 2.5.** *Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha < 2$ . Dann gibt es eine langsam variierende Funktion  $L$ , so dass*

$$1 - \mu(\theta) \sim \frac{\Gamma(2 - \alpha)}{\alpha - 1} (1 - \theta)^{\alpha - 1} L\left(\frac{1}{1 - \theta}\right), \quad \text{für } \theta \uparrow 1 \quad (2.30)$$

und

$$\mu(\theta, i) \sim \Gamma(i - \alpha) (1 - \theta)^{-(i - \alpha)} L\left(\frac{1}{1 - \theta}\right), \quad \text{für } \theta \uparrow 1, i \geq 2. \quad (2.31)$$

**Beweis.** Wir beweisen zuerst (2.31). Sei nun  $i \geq 2$  und  $X$  eine Zufallsvariable mit Verteilung  $(p_j)_{j \geq 0}$ . Nach Feller [13], Kapitel XVII.5, kann man das asymptotische Verhalten des gestutzten  $i$ -ten Moments von  $X$  durch das Verhalten der Schwanzverteilung von  $X$  beschreiben,

$$\mathbb{E}[X^i; X < x] \sim \frac{\alpha}{i - \alpha} x^i \mathbb{P}(X > x) \quad \text{für } x \rightarrow \infty.$$

Nach Korollar 2 in Kapitel XVII.5 [13], variiert der Schwanz der Verteilung  $\mathcal{L}(X)$  regulär mit Exponent  $-\alpha$ , d.h.

$$\mathbb{P}(X > x) \sim x^{-\alpha} \bar{L}(x), \quad \text{für } x \rightarrow \infty,$$

mit einer langsam variierenden Funktion  $\bar{L}$ . Damit erhält man

$$\sum_{j=i}^{\lfloor x \rfloor} [j]_i p_j \sim \mathbb{E}[X^i; X < x] \sim \frac{\alpha}{i - \alpha} x^{i - \alpha} \bar{L}(x), \quad \text{für } x \rightarrow \infty. \quad (2.32)$$

Den Zusammenhang zwischen der Asymptotik des gestutzten  $i$ -ten Moments und der Asymptotik der faktoriellen Momente erhält man mittels eines Taubersatzes. Theorem 5 in Kapitel XIII.5, [13] liefert die Äquivalenz von (2.32) zu

$$\mu(\theta, i) = (f(\theta))^{-1} \sum_{j=i}^{\infty} [j]_i p_j \theta^j \sim \Gamma(i - \alpha) (1 - \theta)^{-(i - \alpha)} L\left(\frac{1}{1 - \theta}\right) \quad \text{für } \theta \uparrow 1,$$

wobei  $L = \alpha \bar{L}$  nicht von  $i$  abhängt. Zum Beweis von (2.30) betrachte für  $\theta \uparrow 1$

$$\begin{aligned} 1 - \mu(\theta) &\sim f(\theta) - \theta f'(\theta) = \int_{\theta}^1 s f''(s) ds \\ &= \int_{(1 - \theta)^{-1}}^{\infty} (1 - x^{-1}) x^{-2} f''(1 - x^{-1}) dx \sim \int_{(1 - \theta)^{-1}}^{\infty} g(x) dx, \end{aligned} \quad (2.33)$$

wobei  $g(x) = x^{-2}f''(1 - x^{-1})$  eine regulär variierende Funktion zum Index  $-\alpha$  ist. Für  $i = 2$  erhält man nach Definition (2.13) und Asymptotik (2.31)

$$g(x) \sim x^{-2}\mu(1 - x^{-1}, 2) \sim \Gamma(2 - \alpha)x^{-\alpha}L(x) \text{ für } x \rightarrow \infty. \quad (2.34)$$

Schließlich liefert eine Anwendung von Karamatas Theorem (vgl. Theorem 1.5.11 in [8]) die Behauptung von Lemma 2.5

$$\begin{aligned} 1 - \mu(\theta) &\sim \int_{(1-\theta)^{-1}}^{\infty} g(x) dx \\ &\sim \frac{1}{(\alpha - 1)(1 - \theta)} g\left(\frac{1}{1 - \theta}\right) \\ &\sim \frac{\Gamma(2 - \alpha)}{\alpha - 1} (1 - \theta)^{\alpha - 1} L\left(\frac{1}{1 - \theta}\right) \text{ für } \theta \uparrow 1. \quad \square \end{aligned}$$

Analog wird der Fall  $\alpha = 2$  behandelt. Dabei können aber nur Aussagen über das asymptotische Verhalten des ersten und zweiten Moments getroffen werden.

**Lemma 2.6.** *Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich der Normalverteilung. Dann gibt es eine langsam variierende Funktion  $L$  so, dass*

$$1 - \mu(\theta) \sim (1 - \theta) L\left(\frac{1}{1 - \theta}\right), \text{ für } \theta \uparrow 1 \quad (2.35)$$

und

$$\mu(\theta, 2) \sim L\left(\frac{1}{1 - \theta}\right), \text{ für } \theta \uparrow 1. \quad (2.36)$$

**Beweis.** Im Fall  $\sigma^2 < \infty$  kann man nach (2.22) die langsam variierende Funktion  $L$  in (2.35) und (2.36) durch  $\sigma^2$  ersetzen. Wir betrachten nun den Fall  $\mathbb{E}X^2 = \infty$  und folgen dem Beweis von Lemma 2.5. Nach Theorem 2.(a) aus Kapitel XVII.5 [13] gilt mit einer langsam variierenden Funktion  $L$

$$\mathbb{E}[X^2; X < x] \sim L(x), \text{ für } x \rightarrow \infty.$$

Eine weitere Anwendung des Taubersatzes für Potenzreihen zeigt das dies äquivalent zu (2.36) ist. Zum Beweis von (2.35) betrachten wir das asymptotische Verhalten von (2.33). Nach (2.13) und (2.36) gilt

$$g(x) \sim x^{-2}\mu(1 - x^{-1}, 2) \sim x^{-2}L(x) \text{ für } x \rightarrow \infty.$$



Die Anwendung des Theorems von Karamata liefert

$$\begin{aligned} 1 - \mu(\theta) &\sim \int_{(1-\theta)^{-1}}^{\infty} g(x) dx \\ &\sim \frac{1}{1-\theta} g\left(\frac{1}{1-\theta}\right) \sim (1-\theta) L\left(\frac{1}{1-\theta}\right) \text{ für } \theta \uparrow 1. \end{aligned}$$

□

**Beweis von Theorem 2.4.** Wir betrachten zuerst den Fall  $1 < \alpha < 2$ .

Sei  $k \geq 2$  und  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k$ . Nach (2.16) und Lemma 2.5, gilt für  $\theta \uparrow 1$

$$\begin{aligned} c_{\theta,k}^{-1} \mathbb{P}_{\theta}(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) &= (1 - \mu(\theta))^{-s(\bar{r})} \prod_{i \geq 1} \left( \frac{\mu(\theta, i)}{i!} \right)^{\beta_i(\bar{r})} \\ &\sim \left( \frac{\Gamma(2-\alpha)}{\alpha-1} (1-\theta)^{\alpha-1} L\left(\frac{1}{1-\theta}\right) \right)^{-s(\bar{r})} \\ &\quad \cdot \prod_{i \geq 2} \left( \frac{\Gamma(i-\alpha)}{\Gamma(i+1)} (1-\theta)^{-(i-\alpha)} L\left(\frac{1}{1-\theta}\right) \right)^{\beta_i(\bar{r})} \\ &= (1-\theta)^{-h_1(\bar{r})} \left( \frac{\Gamma(2-\alpha)}{\alpha-1} L\left(\frac{1}{1-\theta}\right) \right)^{-h_2(\bar{r})} \prod_{i \geq 2} q_i(\alpha)^{\beta_i(\bar{r})}, \quad (2.37) \end{aligned}$$

wobei (siehe (1.7))

$$\begin{aligned} h_1(\bar{r}) &= (\alpha-1)s(\bar{r}) + \sum_{i \geq 2} (i-\alpha)\beta_i(\bar{r}) \\ &= \alpha(\beta_0(\bar{r}) + \beta_1(\bar{r})) - \beta_1(\bar{r}) - 1 \end{aligned} \quad (2.38)$$

und

$$h_2(\bar{r}) = s(\bar{r}) - \sum_{i \geq 2} \beta_i(\bar{r}) = \beta_0(\bar{r}) + \beta_1(\bar{r}).$$

Da jeder Knoten  $v \in \bar{r}$  mit  $c(v, \bar{r}) \in \{0, 1\}$  in der Menge der ausgezeichneten Knoten  $\{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}$  liegt, gilt  $\beta_0(\bar{r}) + \beta_1(\bar{r}) \leq k$ . Damit ist  $h_1(\bar{r})$  maximal (und gleich  $\alpha k - 1$ ) genau dann, wenn  $\beta_0(\bar{r}) = k$ .

Nach Lemma 2 in Kapitel VIII.8 [13] gilt

$$\lim_{\theta \uparrow 1} (1-\theta)^{\varepsilon} L\left(\frac{1}{1-\theta}\right) = 0, \quad \forall \varepsilon > 0.$$

Wir erhalten schließlich (2.27) aus (2.37) mit der Normierung

$$c_k(\alpha) = \sum_{\bar{r} \in \mathcal{R}_k^*} \prod_{j \geq 2} q_j(\alpha)^{\beta_j(\bar{r})}.$$

Im Fall  $\alpha = 2$  und  $\sigma^2 < \infty$  folgt Behauptung (2.29) aus Theorem 2.3. Den Fall  $\alpha = 2$  und  $\sigma^2 = \infty$  beweist man analog des Beweises von Theorem 2.3. Dazu benutzt man folgende Verallgemeinerung von (2.26)

$$\lim_{\theta \uparrow 1} (1 - \theta)^{i-2} \frac{\mu(\theta, i)}{\mu(\theta, 2)} = 0, \quad \text{für } i \geq 3. \quad (2.39)$$

Letzteres erhält man durch Anwendung von Theorem 1.3.3 in [8]. Es bleibt nun noch zu zeigen, dass die Normierungskonstante  $c_k(\alpha)$  durch (2.29) dargestellt werden kann. Dazu definieren wir zunächst die Verteilung  $\mathcal{L}_1$  des reduzierten Skeletts

$$\mathbb{P}_1(\bar{R}_k = (\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\})) := (c_k(\alpha))^{-1} \prod_{j \geq 2} q_j(\alpha)^{\beta_j(\bar{r})} \quad (2.40)$$

für  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k^*$ . Zum Beweis von (2.29) werden wir zeigen, dass die zufälligen minimalen Bäume  $\bar{R}_k$ ,  $k \geq 2$  unter  $\mathcal{L}_1$  konsistent im Sinne von Aldous [3] sind: Das von  $k$  rein zufällig ausgewählten Blättern in  $\bar{R}_{k+1}$  aufgespannte und reduzierte Skelett hat dieselbe Verteilung wie  $\bar{R}_k$ .

Wir benutzen hier wieder die eindeutige Darstellung eines reduzierten  $k$ -Baums  $(\bar{r}, \{\bar{v}^{(1)}, \dots, \bar{v}^{(k)}\}) \in \mathcal{R}_k^*$  durch den Baum  $\bar{r}$ . Sei  $M_k$  die Menge von  $k$  verschiedenen, rein zufällig gezogenen Blättern aus  $\Lambda(\bar{R}_{k+1})$ . Dann gilt für jedes  $t \in \mathcal{R}_k^*$

$$\begin{aligned} & \mathbb{P}_1(r(M_k) \text{ ist äquivalent zu } t) \\ &= \sum_{\bar{r} \in \mathcal{R}_{k+1}^*} \mathbb{P}_1(\bar{R}_{k+1} = \bar{r}) \mathbb{P}_1(r(M_k) \text{ ist äquivalent zu } t \mid \bar{R}_{k+1} = \bar{r}) \\ &= \frac{1}{k+1} \sum_{\bar{r} \in \mathcal{R}_{k+1}^*} \sum_{\substack{w \in \Lambda(\bar{r}): \\ r(\Lambda(\bar{r}) \setminus \{w\}) \simeq t}} \mathbb{P}_1(\bar{R}_k = t) \frac{\mathbb{P}_1(\bar{R}_{k+1} = \bar{r})}{\mathbb{P}_1(\bar{R}_k = t)}. \end{aligned}$$

Zur Berechnung dieser Summe gehen wir nun zu einer äquivalenten Summe über. Wir summieren nun über alle Möglichkeiten, die das reduzierte  $k$ -Skelett  $t$  zu einem Paar  $(\bar{r}, w)$  mit  $w \in \Lambda(\bar{r})$  derart erweitern, dass man durch Entfernen des Blattes  $w$  von  $\bar{r}$  den Baum  $t$  erhält. Dabei werden zwei Fälle unterschieden, nämlich ob das Blatt  $w$  in  $\bar{r}$  eines von mindestens drei oder eines von genau zwei Geschwistern ist. Im ersten Fall bekommt man durch Hinzufügen eines weiteren

Nachfolger  $w$  an einen inneren Knoten  $v$  in  $t$  ein Paar  $(\bar{r}, w)$ . Hat  $v$  genau  $i$  Nachfolger in  $t$ , so gibt es  $i + 1$  Positionen, um  $w$  als neues Blatt einzufügen. In diesem Fall gilt  $\beta_i(t) = \beta_i(\bar{r}) + 1$ ,  $\beta_{i+1}(t) = \beta_{i+1}(\bar{r}) - 1$  und  $\beta_j(t) = \beta_j(\bar{r})$ , für  $j \notin \{0, i, i + 1\}$ . Im zweiten Fall entsteht das Paar  $(\bar{r}, w)$  durch Hinzufügen von zwei neuen Knoten. Zu einem  $v \in t$  wird ein neuer Vorgänger und davon abstammend ein neues Geschwister  $w$  eingefügt. Dabei kann  $w$  links oder rechts von  $v$  liegen. In diesem Fall gilt also  $\beta_2(\bar{r}) = \beta_2(t) + 1$  und  $\beta_j(\bar{r}) = \beta_j(t)$  für alle  $j \notin \{0, 2\}$  (vgl. Abbildung 4.1).

Diese Überlegungen liefern zusammen mit Notation (2.28) die Darstellung

$$\begin{aligned}
 \mathbb{P}_1(r(M_k) \simeq t) &= \frac{1}{k+1} \left( \sum_{v \in t \setminus \Lambda(t)} \mathbb{P}_1(\bar{R}_k = t) (c(v, t) + 1) \frac{c_k(\alpha)}{c_{k+1}(\alpha)} \frac{q_{c(v, t)+1}(\alpha)}{q_{c(v, t)}(\alpha)} \right. \\
 &\quad \left. + \sum_{v \in t} \mathbb{P}_1(\bar{R}_k = t) 2 \frac{c_k(\alpha)}{c_{k+1}(\alpha)} q_2(\alpha) \right) \\
 &= \frac{c_k(\alpha)}{c_{k+1}(\alpha)} \frac{\mathbb{P}_1(\bar{R}_k = t)}{k+1} \left( \sum_{v \in t \setminus \Lambda(t)} (c(v, t) - \alpha) + \sum_{v \in t} (\alpha - 1) \right) \\
 &= \frac{c_k(\alpha)}{c_{k+1}(\alpha)} \frac{\alpha k - 1}{k+1} \mathbb{P}_1(\bar{R}_k = t). \tag{2.41}
 \end{aligned}$$

Für die letzte Gleichheit wurde (1.7) und  $\beta_0(t) = k$  für  $t \in \mathcal{R}_k^*$  verwendet. Summiert man (2.41) über alle  $t \in \mathcal{R}_k^*$ , so erhält man die Rekursion

$$c_{k+1}(\alpha) = \frac{\alpha k - 1}{k+1} c_k(\alpha), \quad k \geq 2.$$

Mit  $c_2(\alpha) = q_2(\alpha) = \frac{\alpha-1}{2}$  folgt schließlich die Behauptung (2.29).  $\square$

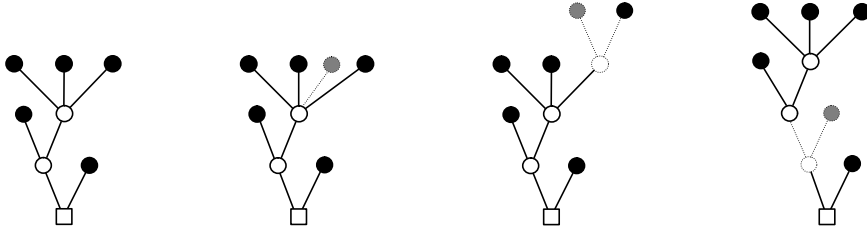


ABBILDUNG 4.1: Einige Beispiele zur Erweiterung eines reduzierten  $k$ -Skeletts  $t$  zu einem Paar  $(\bar{r}, w)$ . Links das reduzierte 5-Skelett  $t \in \mathcal{R}_5^*$ . Rechts davon eine Erweiterung von  $t$  indem ein Blatt  $w$  (grau schraffiert) an einem inneren Knoten von  $t$  eingefügt wird. In den beiden nächsten Erweiterungen wurde ein neuer Vorgänger zu einem Knoten  $v \in t$  mit  $w$  als zweitem Kind eingefügt, wobei zuerst  $v$  ein Blatt und danach  $v$  ein innerer Knoten von  $t$  ist.

## 2.4 Größe des größenverzerrten Baumes

In diesem Kapitel beweisen wir einen Grenzwertsatz für die Verteilung der Größe des größenverzerrten Baumes im Limes  $\theta \uparrow 1$ . Dazu nutzen wir die probabilistische Konstruktion von  $\hat{T}$ , um eine Darstellung von  $s(\hat{T})$  als Summe unabhängiger Zufallsvariablen zu erhalten.

### Darstellung als Summe unabhängiger Zufallsvariablen

Nach Konstruktion 2.1 setzt sich die Anzahl der Knoten des Baumes  $\hat{T}$  aus der Anzahl der Knoten des aufgedickten Skeletts und der Anzahl der Knoten der angefügten, unabhängigen GW-Bäume zusammen. Die Verteilung  $\mathcal{L}_{\theta,k}(s(\hat{T}))$  lässt sich somit nach Proposition 2.2 darstellen durch

$$\mathcal{L}_{\theta,k}(s(\hat{T})) = \mathcal{L}_{\theta}\left(V_k + \sum_{m=1}^{Y_k} s(T_m)\right) \text{ für } 0 < \theta < 1 \text{ und } k \geq 1. \quad (2.42)$$

Dabei sind die Zufallsvariablen  $T_m$ ,  $m \geq 1$  unter  $\mathcal{L}_{\theta}$  unabhängige GW-Bäume mit Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  und unabhängig von  $(V_k, Y_k)$ .  $V_k$  sei die Anzahl der inneren Knoten und  $Y_k$  die Anzahl der Blätter des aufgedickten Skeletts aus Schritt 1-3 der Konstruktion. Das Tupel  $(V_k, Y_k)$  lässt sich nach Konstruktion weiter zerlegen in

$$V_k = s(\bar{R}_k) - \beta_0(\bar{R}_k) + \sum_{j=1}^{s(\bar{R}_k)} L_j, \quad (2.43)$$

$$Y_k = \beta_0(\bar{R}_k) + \sum_{m=1}^{L_1 + \dots + L_{s(\bar{R}_k)}} (\hat{X}_m^{(1)} - 1) + \sum_{i \geq 1} \sum_{m=1}^{\beta_i(\bar{R}_k)} (\hat{X}_m^{(i)} - i). \quad (2.44)$$

Die Zufallsvariablen in (2.43) und (2.44) sind unter  $\mathcal{L}_{\theta}$  wie folgt verteilt:

- $\bar{R}_k$  hat die Verteilung (2.16).
- $L_j$ ,  $j \geq 1$  sind geometrisch zum Parameter  $1 - \mu(\theta)$  verteilt.
- Für jedes  $i \geq 1$  besitzt  $\hat{X}_m^{(i)}$ ,  $m \geq 1$  eine  $i$ -fach größenverzerrte Kinderzahlverteilung  $(p_j(\theta, i))_{j \geq 0}$ .
- Die Zufallsvariablen  $\bar{R}_k$ ,  $L_j$ ,  $j \geq 1$  und  $\hat{X}_m^{(i)}$ ,  $m, i \geq 1$  sind unabhängig.

Darstellung (2.43) setzt sich zusammen aus der Anzahl der inneren Knoten des reduzierten  $k$ -Skeletts  $\bar{R}_k$  aus Schritt 1 und der Anzahl der Knoten, die

durch Einfügen von Knotenketten in Schritt 2 der Konstruktion hinzukommen. Darstellung (2.44) summiert die Anzahl der Blätter von  $\bar{R}_k$  mit den neuen Kindern von Knoten aus Schritt 2 und den neuen Kinder der inneren Knoten von  $\bar{R}_k$ .

### Limes der Verteilungen

Wir beginnen mit einem Grenzwertsatz über die Verteilung der Anzahl der Knoten des aufgedickten Skeletts. Mit  $\text{Gamma}(\beta, \gamma)$  bezeichnen wir die Gammaverteilung zum Skalen- und Formparameter  $\beta, \gamma > 0$ . Die Gammaverteilung  $\text{Gamma}(\beta, \gamma)$  besitzt die Dichte

$$g_{\beta, \gamma}(x) = \frac{\beta^\gamma}{\Gamma(\gamma)} x^{\gamma-1} \exp(-\beta x), \quad x \geq 0$$

und Laplace-Transformierte

$$\varphi_{\beta, \gamma}(\lambda) := \int_0^\infty e^{-\lambda x} g_{\beta, \gamma}(x) dx = \left(1 + \frac{\lambda}{\beta}\right)^{-\gamma}, \quad \lambda \geq 0. \quad (2.45)$$

Da die Laplace-Transformierte einer Summe unabhängiger Zufallsvariablen das Produkt der einzelnen Transformaten ist, erkennt man aus (2.45) sofort, dass die Summe unabhängiger  $\text{Gamma}(\beta, \gamma_1)$  und  $\text{Gamma}(\beta, \gamma_2)$  verteilter Zufallsvariablen  $\text{Gamma}(\beta, \gamma_1 + \gamma_2)$  verteilt ist. Im Folgenden werden wir diese Eigenschaft ohne es weiter zu erwähnen mehrfach benutzen.

**Proposition 2.7.** *Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann folgt für  $k \geq 1$*

$$\mathcal{L}_\theta((1 - \theta)Y_k) \xrightarrow{w} \text{Gamma}(1, \alpha k - 1), \quad \text{für } \theta \uparrow 1. \quad (2.46)$$

**Beweis.** Mit Hilfe der Darstellung (2.44) zeigen wir, dass

$$\lim_{\theta \uparrow 1} \mathbb{E}_\theta \exp(-\lambda(1 - \theta)Y_k) = (1 + \lambda)^{-(\alpha k - 1)}, \quad \lambda \geq 0.$$

Behauptung (2.46) folgt dann aus der Eindeutigkeit der Laplace-Transformierten. Wir beweisen zuerst, dass für  $1 < \alpha < 2$  und  $i \geq 2$ ,

$$\mathcal{L}_\theta((1 - \theta)\hat{X}_1^{(i)}) \xrightarrow{w} \text{Gamma}(1, i - \alpha), \quad \text{für } \theta \uparrow 1 \quad (2.47)$$

und für  $\alpha = 2$

$$\mathcal{L}_\theta((1 - \theta)\hat{X}_1^{(2)}) \xrightarrow{w} \delta_0, \quad \text{für } \theta \uparrow 1. \quad (2.48)$$

Dazu berechnen wir die Erzeugendenfunktion von  $\hat{X}_1^{(i)} - i$ . Mit (2.2), (2.12) und (2.13) sieht man, dass für jede beliebige Kinderzahlverteilung  $(p_j)_{j \geq 0}$ , jedes  $i \geq 1$  und  $0 < \theta < 1$

$$\begin{aligned} \mathbb{E}_\theta s^{\hat{X}_1^{(i)} - i} &= \sum_{j \geq i} s^{j-i} p_j(\theta, i) \\ &= \frac{f^{(i)}(\theta s)}{f^{(i)}(\theta)} = \frac{f(\theta s)}{s^i f(\theta)} \frac{\mu(\theta s, i)}{\mu(\theta, i)}, \quad 0 \leq s \leq 1. \end{aligned} \quad (2.49)$$

Damit erhält man für alle  $\lambda \geq 0$

$$\mathbb{E}_\theta \exp(-\lambda(1-\theta)\hat{X}_1^{(i)}) \sim \frac{\mu(\theta e^{-\lambda(1-\theta)}, i)}{\mu(\theta, i)}, \quad \text{für } \theta \uparrow 1.$$

Nach (2.31), bzw. (2.36) und der Eigenschaft der langsam variierenden Funktion  $L$  (für alle  $c > 0$  gilt  $L(cx)/L(x) \rightarrow 1$  mit  $x \rightarrow \infty$ ), gilt

$$\lim_{\theta \uparrow 1} \mathbb{E}_\theta \exp(-\lambda(1-\theta)\hat{X}_1^{(i)}) = \begin{cases} (1+\lambda)^{-(i-\alpha)}, & \text{für } 1 < \alpha < 2 \text{ und } i \geq 2, \\ 1, & \text{für } \alpha = 2 \text{ und } i = 2. \end{cases}$$

Dies sind die Laplace-Transformierten der Verteilungen in (2.47) und (2.48). Im nächsten Schritt zeigen wir für  $1 < \alpha \leq 2$  und  $n \geq 1$

$$\mathcal{L}_\theta \left( (1-\theta) \sum_{m=1}^{L_1+\dots+L_n} (\hat{X}_m^{(1)} - 1) \right) \xrightarrow{w} \text{Gamma}(1, (\alpha-1)n), \quad \theta \uparrow 1. \quad (2.50)$$

Zum Beweis von (2.50) genügt es, aufgrund der Unabhängigkeit der Zufallsvariablen, den Fall  $n = 1$  zu betrachten. Die Erzeugendenfunktion von  $L_1$  (vgl. (2.4)) ist

$$\mathbb{E}_\theta s^{L_1} = \frac{1 - \mu(\theta)}{1 - \mu(\theta)s} = \frac{f(\theta) - \theta f'(\theta)}{f(\theta) - \theta f'(\theta)s}, \quad 0 \leq s \leq 1. \quad (2.51)$$

Mittels (2.49) und (2.51) erhält man

$$\begin{aligned} \mathbb{E}_\theta \exp \left( -\lambda(1-\theta) \sum_{m=1}^{L_1} (\hat{X}_m^{(1)} - 1) \right) \\ = \mathbb{E}_\theta \left( \frac{f'(u_\theta)}{f'(\theta)} \right)^{L_1} = \frac{f(\theta) - \theta f'(\theta)}{f(\theta) - \theta f'(u_\theta)}, \end{aligned} \quad (2.52)$$

wobei zur besseren Übersicht  $u_\theta := \theta e^{-\lambda(1-\theta)}$  gesetzt wurde. Um nun den Limes  $\theta \uparrow 1$  zu bilden, betrachten wir das asymptotische Verhalten von Zähler und Nenner in (2.52) zuerst getrennt voneinander. Mit (2.30) und (2.35) folgt

$$f(\theta) - \theta f'(\theta) = f(\theta)(1 - \mu(\theta)) \sim c'_\alpha (1 - \theta)^{\alpha-1} L\left(\frac{1}{1-\theta}\right) \quad \text{für } \theta \uparrow 1, \quad (2.53)$$

mit  $c'_\alpha := \Gamma(2 - \alpha)/(\alpha - 1)$  für  $1 < \alpha < 2$  und  $c'_2 := 1$ . Ersetzt man  $\theta$  in (2.53) durch  $u_\theta$ , so folgt mit Hilfe eines Mittelwertsatzes und den Asymptotiken (2.31) und (2.36)

$$\begin{aligned} f(\theta) - \theta f'(u_\theta) &= f(u_\theta) - u_\theta f'(u_\theta) \\ &\quad + f(\theta) - f(u_\theta) - (\theta - u_\theta) f'(u_\theta) \\ &\sim c'_\alpha (1 - u_\theta)^{\alpha-1} L\left(\frac{1}{1 - u_\theta}\right) \text{ für } \theta \uparrow 1. \end{aligned} \quad (2.54)$$

Die Unabhängigkeit der Zufallsvariablen  $L_j$  kombiniert mit (2.52), (2.53) und (2.54) liefert die Laplace-Transformierte der Gammaverteilung in (2.50)

$$\lim_{\theta \uparrow 1} \mathbb{E}_\theta \exp\left(-\lambda(1 - \theta) \sum_{m=1}^{L_1} (\hat{X}_m^{(1)} - 1)\right) = (1 + \lambda)^{-(\alpha-1)}.$$

Sei nun  $k \geq 1$  und  $\bar{r} \in \mathcal{R}_k^*$ . Für  $1 < \alpha < 2$  folgt aus der Darstellung von  $Y_k$  in (2.44) und den Limiten (2.47) und (2.50)

$$\mathcal{L}_\theta((1 - \theta)Y_k \mid \bar{R}_k = \bar{r}) \xrightarrow{w} \text{Gamma}(1, h_1(\bar{r})), \text{ für } \theta \uparrow 1, \quad (2.55)$$

wobei  $h_1$  die Funktion aus (2.38) ist. Aus der zweiten Gleichung in (2.38) folgt, dass  $h_1(\bar{r}) = \alpha k - 1$  für  $\bar{r} \in \mathcal{R}_k^*$ . Da die Grenzverteilung (2.55) unabhängig von  $\bar{r} \in \mathcal{R}_k^*$  ist und  $\lim_{\theta \uparrow 1} \mathbb{P}_\theta(\bar{R}_k \in \mathcal{R}_k^*) = 1$ , folgt somit (mit Theorem 2.4) die Behauptung der Proposition für  $1 < \alpha < 2$ .

Sei jetzt  $\alpha = 2$  und  $\bar{r} \in \mathcal{R}_k^*$  ein vollständig binärer Baum, d.h.  $\beta_2(\bar{r}) = k - 1$  und  $\beta_i(\bar{r}) = 0$  für  $i \geq 3$ . Hier liefert die Darstellung von  $Y_k$  in (2.44) und die Limiten (2.48) und (2.50) die schwache Konvergenz

$$\mathcal{L}_\theta((1 - \theta)Y_k \mid \bar{R}_k = \bar{r}) \xrightarrow{w} \text{Gamma}(1, 2k - 1), \theta \uparrow 1.$$

Nach Theorem 2.4 gilt in diesem Fall  $\mathbb{P}_\theta(\bar{R}_k \in \mathcal{R}_k^*, \beta_2(\bar{R}_k) = k - 1) \rightarrow 1$  für  $\theta \uparrow 1$ .

□

Wir kommen nun zum Hauptresultat dieses Kapitels, einem Theorem über die Grenzverteilung der reskalierten Größe des Baumes  $\hat{T}$ . Wir beweisen dieses Theorem unter einer etwas stärkeren Bedingung an die Kinderzahlverteilung und setzen voraus, dass  $(p_j)_{j \geq 0}$  im normalen Anziehungsbereich einer stabilen Verteilung ist. In diesem Fall gilt für die Erzeugendenfunktion

$$f(\theta) - \theta \sim C_\alpha (1 - \theta)^\alpha \text{ für } \theta \uparrow 1 \quad (2.56)$$

mit einer Konstanten  $C_\alpha > 0$ . Die langsam variierende Funktion in Lemma 2.5 und 2.6 kann hier durch die Konstante  $\alpha C_\alpha$  ersetzt werden (siehe Kapitel XVII.5 in [13]).

**Theorem 2.8.** *Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im normalen Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt für  $k \geq 1$ ,*

$$\mathcal{L}_{\theta,k}((f(\theta) - \theta)s(\hat{T})) \xrightarrow{w} \text{Gamma}(1, k - \frac{1}{\alpha}), \quad \theta \uparrow 1. \quad (2.57)$$

**Bemerkungen.**

**Reskalierung.** Die Reskalierung in (2.57) entspricht asymptotisch, bis auf eine von  $\alpha$  und  $k$  abhängige Konstante, der “natürlichen” Skalierung mit dem Erwartungswert  $\mathbb{E}_{\theta,k}s(\hat{T})$ . Nach (2.18) gilt nämlich

$$\mathbb{E}_{\theta,k}s(\hat{T}) = \frac{\mathbb{E}_{\theta,0}[s(\hat{T})]_{k+1}}{\mathbb{E}_{\theta,0}[s(\hat{T})]_k} + k.$$

Die Asymptotik von  $\mathbb{E}_{\theta,0}[s(\hat{T})]_k$  für  $\theta \uparrow 1$  ergibt sich aus (2.29), (2.37) und (2.56)

$$\mathbb{E}_{\theta,0}[s(\hat{T})]_k \sim \left( \frac{\Gamma(2-\alpha)}{\alpha-1} \alpha C_\alpha \right)^{-k} c_k(\alpha) (1-\theta)^{-\alpha k+1}, \quad \text{für } \theta \uparrow 1.$$

Schließlich erhält man

$$\mathbb{E}_{\theta,k}s(\hat{T}) \sim d_{\alpha,k} (C_\alpha(1-\theta)^\alpha)^{-1} \sim d_{\alpha,k} (f(\theta) - \theta)^{-1},$$

wobei

$$d_{\alpha,k} := \frac{(k\alpha - 1)(\alpha - 1)}{(k + 1)\alpha\Gamma(2 - \alpha)}.$$

**Höhe eines uniform gezogenen Knotens.** Nach dem Satz der totalen Wahrscheinlichkeit lässt sich die Verteilung der Höhe eines uniform aus  $\hat{T}$  gezogenen Knotens darstellen durch

$$\mathbb{P}_{\theta,k}(|B_1| = h) = \sum_{n \geq h} \mathbb{P}_{\theta,k}(|B_1| = h \mid s(\hat{T}) = n) \mathbb{P}_{\theta,k}(s(\hat{T}) = n).$$

Die Verteilung der Höhe des Knotens  $B_1$  bedingt auf die Gesamtgröße des Baumes ist aber von den Parametern  $\theta$  und  $k$  unabhängig. Im reskalierten Limes der einzelnen Verteilungen erhält man für die Verteilung der Höhe von  $B_1$  eine Exponentialverteilung sowie für die Verteilung der Größe von  $\hat{T}$  nach obigen Satz eine Gammaverteilung. Im Fall  $\alpha = 2$  ist dies konsistent mit dem Resultat von Aldous über die Höhe eines uniform gezogenen Knotens im bedingten



Baum. Integriert man die Dichte  $g$  aus (2.17) mit der Gammaverteilung zum Formparameter  $\frac{1}{2}$  (dies entspricht dem Fall  $k = 1$  und  $\alpha = 2$ ) so erhält man eine Exponentialverteilung. Dies entspricht der reskalierten Verteilung der Höhe eines uniform aus  $\hat{T}$  unter  $\mathcal{L}_{\theta,1}$  gezogenen Knotens für  $\theta \uparrow 1$  (vergleiche die Bemerkung auf Seite 25). Im Fall binärer Verzweigung ist das Resultat von Theorem 2.8 implizit in [18] enthalten.

**Beweis.** Sei  $k \geq 1$ . Wir benutzen wieder die Darstellung von  $\mathcal{L}_{\theta,k}(s(\hat{T}))$  in (2.42). Da die Zufallsvariablen  $L_j$  in (2.43) unabhängig, geometrisch verteilt zum Parameter  $1 - \mu(\theta)$  und unabhängig von  $\bar{R}_k$  sind, folgt mit den Asymptotiken (2.30) (bzw. (2.35), für  $\alpha = 2$ ) und (2.56)

$$(f(\theta) - \theta) \mathbb{E}_\theta V_k = (f(\theta) - \theta) \left( \frac{\mu(\theta)}{1 - \mu(\theta)} \mathbb{E}_\theta s(\bar{R}_k) + O(1) \right) = O(1 - \theta) \text{ für } \theta \uparrow 1.$$

Damit gilt für alle  $\lambda > 0$

$$\mathbb{E}_{\theta,k} \exp \left( -\lambda (f(\theta) - \theta) s(\hat{T}) \right) \sim \mathbb{E}_\theta \exp \left( -\lambda (f(\theta) - \theta) \sum_{m=1}^{Y_k} s(T_m) \right) \text{ für } \theta \uparrow 1. \quad (2.58)$$

Zur Bestimmung des Limes in (2.58) nutzen wir folgende Beziehung zwischen der Größe eines GW-Baumes  $T$  und der ersten Trefferzeit einer assoziierten Irrfahrt. Unabhängig der Kinderzahlverteilung  $(p_j)_{j \geq 0}$  gilt folgende Gleichheit in Verteilung (siehe [12])

$$s(T) \stackrel{d}{=} \min\{n \geq 0 : S_n = -1\}, \quad (2.59)$$

wobei  $(S_n)_{n \geq 0}$  ein linksstetige Irrfahrt ist, die in 0 startet und deren Zuwächse gemäß  $\mathbb{P}(S_{n+1} - S_n = j) = p_{j+1}$ ,  $j \geq -1$  verteilt sind. (Zur Bijektion zwischen Irrfahrtsexkursionen und verwurzelten planaren Bäumen und zum Beweis der Identität (2.59) siehe [6, 29].) Mit (2.59) und der Linksstetigkeit von  $(S_n)_{n \geq 0}$  sieht man nun, dass mit unabhängigen Kopien  $T_m$ ,  $m \geq 1$  von  $T$  für alle  $\ell \geq 1$

$$\sum_{m=1}^{\ell} s(T_m) \stackrel{d}{=} \eta_\ell, \quad (2.60)$$

wobei  $\eta_\ell := \min\{n \geq 0 : S_n \leq -\ell\}$  die erste Trefferzeit ist, zu der die Irrfahrt  $(S_n)_{n \geq 0}$  in das Intervall  $(-\infty, -\ell]$  eintritt. Ist nun die Kinderzahlverteilung  $(p_j)_{j \geq 0}$  im normalen Anziehungsbereich einer stabilen Verteilung, so befinden sich die Trefferzeiten  $\eta_\ell$  im Anziehungsbereich eines stabilen Subordinators mit

Index  $\frac{1}{\alpha}$  (siehe z.B. Theorem 8.9.12 in [8]), d.h. für jedes  $y \geq 0$  gilt

$$C_\alpha \ell^{-\alpha} \eta_{\lfloor y\ell \rfloor} \xrightarrow{d} \tau_y, \text{ für } \ell \rightarrow \infty. \quad (2.61)$$

$\tau_y$  ist die Trefferzeit (First Passage Time) des Intervalls  $(-\infty, -y]$  eines spektral-positiven Lévy Prozesses mit Index  $\alpha$  und Skalenparameter  $C_\alpha$ . Dieser Lévy Prozess ist der Verteilungslimes der reskalierten Irrfahrt  $(n^{-\frac{1}{\alpha}} S_{\lfloor nt \rfloor})_{t \geq 0}$ . Die Laplace-Transformierte von  $\tau_y$  ist

$$\mathbb{E} \exp(-\lambda \tau_y) = \exp\left(-y \lambda^{\frac{1}{\alpha}}\right), \quad \lambda \geq 0. \quad (2.62)$$

Die GW-Bäume  $T_m$ , die wir betrachten, besitzen aber eine subkritische Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$  mit Parameter  $\theta$  anstatt der kritischen Verteilung  $(p_j)_{j \geq 0}$ . Wir zeigen, dass dies einen “Tilt” der Grenzverteilung bewirkt, d.h.

$$\mathcal{L}_\theta \left( (f(\theta) - \theta) \sum_{m=1}^{y(1-\theta)^{-1}} s(T_m) \right) \xrightarrow{w} \mathcal{L}(\tau_y^*) \text{ für } \theta \uparrow 1, \quad (2.63)$$

wobei  $\tau_y^*$  die Dichte

$$f_y^*(x) = \exp(y-x) f_y(x), \quad x \geq 0 \quad (2.64)$$

( $f_y$  sei die Dichte von  $\mathcal{L}(\tau_y)$ ) und die Laplace-Transformierte

$$\mathbb{E} \exp(-\lambda \tau_y^*) = \exp\left(-y \left((1+\lambda)^{\frac{1}{\alpha}} - 1\right)\right), \quad \lambda \geq 0 \quad (2.65)$$

besitzt. Zum Beweis von (2.64) stellen wir mit (2.8) fest, dass

$$\mathbb{P}_\theta \left( \sum_{m=1}^{\ell} s(T_m) = n \right) = \frac{\theta^{n-\ell}}{f(\theta)^n} \mathbb{P} \left( \sum_{m=1}^{\ell} s(T_m) = n \right) \text{ für alle } \ell, n \geq 1.$$

Mit (2.56), (2.60) und (2.61) gilt somit

$$\begin{aligned} & \mathbb{P}_\theta \left( (f(\theta) - \theta) \sum_{m=1}^{y(1-\theta)^{-1}} s(T_m) \in dx \right) \\ & \sim \left( 1 - \frac{f(\theta) - \theta}{f(\theta)} \right)^{x(f(\theta) - \theta)^{-1}} \theta^{-y(1-\theta)^{-1}} \mathbb{P}(C_\alpha (1-\theta)^\alpha \eta_{\lfloor y(1-\theta)^{-1} \rfloor} \in dx) \\ & \longrightarrow \exp(y-x) f_y(x) dx \text{ für } \theta \uparrow 1. \end{aligned}$$

Behauptung (2.65) folgt nun unmittelbar aus (2.62) und (2.64). Theorem 2.8 erhält man durch Anwendung von (2.58), (2.63), Proposition 2.7 und schließlich (2.65)

$$\begin{aligned}
& \mathbb{E}_{\theta,k} \exp \left( -\lambda (f(\theta) - \theta) s(\hat{T}) \right) \\
& \longrightarrow \int_0^\infty \mathbb{E} \exp(-\lambda \tau_y^*) g_{1,\alpha k-1}(y) dy \\
& = \frac{1}{\Gamma(\alpha k - 1)} \int_0^\infty y^{\alpha k-2} \exp \left( - (1 + \lambda)^{\frac{1}{\alpha}} y \right) dy \\
& = (1 + \lambda)^{-(k - \frac{1}{\alpha})} \quad \text{für } \lambda \geq 0. \quad \square
\end{aligned}$$

**Bemerkung.** Theorem 2.8 lässt sich auch auf direktem (aber weniger anschaulichem) Weg beweisen. Man nutzt dazu folgende Variante der Identität (2.59) (siehe [29])

$$\mathbb{P}(s(T) = n) = \frac{1}{n} \mathbb{P}(S_n = -1), \quad n \geq 1.$$

Mit Hilfe von Beziehung (2.8) kann man (2.57) durch Anwendung eines lokalen Grenzwertsatzes über die Irrfahrt  $(S_n)_{n \geq 0}$  herleiten.

## 2.5 Größenverzerrung in $n$ -ter Generation

In diesem Abschnitt stellen wir eine Familie von Verteilungen  $\mathcal{L}_{k,n}$ ,  $k, n \in \mathbb{N}$  eines zufälligen Stammbaumes  $\hat{T}$  vor, die wir durch einen Maßwechsel aus dem unverzerrten GW-Maß  $\mathcal{L}$  erhalten. Dieser Maßwechsel entsteht durch eine Größenverzerrung mit der Größe der  $n$ -ten Generation.

Wir betrachten im folgenden eine kritische Kinderzahlverteilung  $(p_j)_{j \geq 0}$  mit endlichen Momenten beliebiger Ordnung. Mit dieser Voraussetzung sind auch die faktoriellen Momente der Größe der  $n$ -ten Generation  $\mathbb{E}[z_n(T)]_k = f_{(n)}^{(k)}(1)$  für alle  $k, n \geq 0$  endlich.  $f_{(n)}^{(k)}(s)$  ist die  $k$ -te Ableitung der  $n$ -fach iterierten Erzeugendenfunktion  $f$ .

Die Verteilung  $\mathcal{L}_{k,n}$  definieren wir nun für beliebiges  $k, n \geq 1$  als *Größenverzerrung* der Verteilung  $\mathcal{L}$  mit der Größe der  $n$ -ten Generation des Baumes, d.h. für einen endlichen Baum  $t$  mit  $|z_n(t)| \geq k$  sei

$$\mathbb{P}_{k,n}(\hat{T} = t) := \frac{\binom{|z_n(t)|}{k} \mathbb{P}(T = t)}{\mathbb{E} \left( \binom{|z_n(T)|}{k} \right)}. \quad (2.66)$$

Die folgende Betrachtung des Paares  $(\hat{T}, B_k)$ , wobei  $B_k = \{V^{(1)}, \dots, V^{(k)}\}$  eine Menge von  $k$  verschiedenen, rein zufällig aus  $z_n(\hat{T})$  gezogenen Knoten ist, liefert analog zu Kapitel 2.2 eine anschauliche mehrstufige Konstruktion des Baumes  $\hat{T}$  entlang des durch  $B_k$  aufgespannten Skeletts. Die Verteilung von  $\hat{T}$  erhält man dann als Randverteilung dieser multivariaten Zufallsvariable auf dem Raum der Bäume. Es gilt für einen endlichen Baum  $t$  mit  $|z_n(t)| \geq k$  und  $\{v^{(1)}, \dots, v^{(k)}\} \in z_n(t)$

$$\mathbb{P}_{k,n}(\hat{T} = t, B_k = \{v^{(1)}, \dots, v^{(k)}\}) = \frac{k!}{\mathbb{E}[|z_n(T)|]_k} \mathbb{P}(T = t). \quad (2.67)$$

Motiviert durch Konstruktion 2.1 stellen wir (2.67) durch ein Produkt dar, das wir aus folgender vierstufiger Zerlegung des Paares  $(t, \{v^{(1)}, \dots, v^{(k)}\})$  gewinnen: Zerlege  $(t, \{v^{(1)}, \dots, v^{(k)}\})$  in

- (i) das zu  $\{v^{(1)}, \dots, v^{(k)}\}$  äquivalente reduzierte  $k$ -Skelett  $\bar{r}$ ,
- (ii) das durch die Knoten  $v^{(i)}$  aufgespannte Skelett  $r$ ,
- (iii) das durch Hinzunahme der Geschwister aufgedickte Skelett und
- (iv) die restlichen Knoten die es zu  $t$  ergänzen.

Dies liefert mit der größenverzerrten Kinderzahlverteilung  $(p_j(1, i))_{j \geq 0}$  in (2.12) (mit  $\theta = 1$ ) die Darstellung

$$\begin{aligned} \mathbb{P}_{k,n}(\hat{T} = t, B_k = \{v^{(1)}, \dots, v^{(k)}\}) &= \frac{k! c_n(\bar{r}) \prod_{v \in \bar{r}} \frac{f^{(c(v, \bar{r}))}(1)}{c(v, \bar{r})!}}{\mathbb{E}[|z_n(T)|]_k} \\ &\cdot \frac{1}{c_n(\bar{r})} \\ &\cdot \prod_{v \in r} p_{c(v, t)}(1, c(v, r)) \left( \frac{c(v, t)}{c(v, r)} \right)^{-1} \\ &\cdot \prod_{v \in t \setminus r} p_{c(v, t)}. \end{aligned} \quad (2.68)$$

Hier ist  $r := r(v^{(1)}, \dots, v^{(k)})$  und  $\bar{r}$  das zu  $r$  äquivalente reduzierte  $k$ -Skelett.  $c_n(\bar{r})$  ist die Anzahl der zum reduzierten  $k$ -Skelett  $\bar{r} \in \mathcal{R}_k^*$  äquivalenten Bäume  $t$  aus der Menge der Bäume  $\mathcal{R}_k^n$  mit genau  $k$  Blättern in der  $n$ -ten Generation.

$$c_n(\bar{r}) := |\{t \in \mathcal{R}_k^n : t \text{ äquivalent } \bar{r}\}|$$

mit

$$\mathcal{R}_k^n := \{t \subset \mathcal{V} : t \text{ ist ein Baum und } |z_n(t)| = |\Lambda(t)| = k\}.$$

Da hier die ausgezeichneten Knoten ausschließlich Blätter sind, benutzen wir den Begriff der Äquivalenz von Skeletten im Sinne der Äquivalenz der Menge der Blätter (vgl. Bemerkung nach 1.11).

Identität (2.68) liefert folgende vierstufige Konstruktion einer multivariaten Zufallsvariablen mit Verteilung (2.67).

**Konstruktion 2.9.** Sei  $k, n \geq 1$ .

**1. Reduziertes  $k$ -Skelett.** Ziehe ein reduziertes  $k$ -Skelett  $\bar{R}_k$  mit Verteilung

$$\mathbb{P}_n(\bar{R}_k = \bar{r}) := \begin{cases} \frac{k! c_n(\bar{r}) \prod_{v \in \bar{r}} \frac{f^{(c(v, \bar{r}))}(1)}{c(v, \bar{r})!}}{\mathbb{E}[|z_n(T)|]_k}, & \text{für } \bar{r} \in \mathcal{R}_k^*, \\ 0, & \text{sonst.} \end{cases} \quad (2.69)$$

**2.  $k$ -Skelett der Höhe  $n$ .** Gegeben  $\bar{R}_k = \bar{r}$ , ziehe das Skelett  $R_k^n$  uniform aus allen zu  $\bar{r}$  äquivalenten Bäumen mit genau  $k$  Blättern in  $n$ -ter Generation,

$$\mathbb{P}_n(R_k^n = r \mid \bar{R}_k = \bar{r}) := \begin{cases} \frac{1}{c_n(\bar{r})}, & \text{für } r \in \mathcal{R}_k^n \text{ mit } r \text{ äquivalent } \bar{r} \\ 0, & \text{sonst.} \end{cases} \quad (2.70)$$

$B_k^*$  sei die Menge der  $k$  Blätter.

3. **Aufgedicktes  $k$ -Skelett der Höhe  $n$ .** Gegeben  $R_k^n = r$ , ziehe für jeden Knoten  $v \in r \setminus \Lambda(r)$  Kinderzahlen  $X_v$  gemäß einer  $c(v, r)$ -fach größenverzerrten Kinderzahlverteilung und gebe den  $X_v - c(v, r)$  neuen Kindern rein zufällig eine der  $\binom{X_v}{c(v, r)}$  möglichen Anordnungen.
4. **Aufforsten des aufgedickten  $k$ -Skeletts der Höhe  $n$ .** Füge an jedes Blatt des aufgedickten Skeletts einen GW-Baum mit Kinderzahlverteilung  $(p_j)_{j \geq 0}$  an.

Das durch die Konstruktion gewonnene Tupel notieren wir mit  $(T_k^*, B_k^*)$  und folgern unmittelbar aus (2.68)

**Proposition 2.10.** Für jedes  $k, n \geq 1$  gilt

$$\mathcal{L}_n(T_k^*, B_k^*) = \mathcal{L}_{k,n}(\hat{T}, B_k).$$

Wir interessieren uns nun für die Struktur des zufälligen  $k$ -Skeletts  $\bar{R}_k$  sowie die Höhe des MRCA  $|\wedge(B_k^*)|$ . Zunächst betrachten wir die Höhe des MRCA zweier uniform gezogener Knoten. Für  $k = 2$  existiert nur ein reduziertes 2-Skelett  $\bar{r}$ , welches aus einer Wurzel mit zwei Kindern besteht. Es gibt genau  $n$  zu  $\bar{r}$  äquivalente Skelette  $r \in \mathcal{R}_2^n$ , da das 2-Skelette der Höhe  $n$  durch die Höhe des MRCA der 2 Blätter eindeutig festgelegt ist. Damit ist

$$\mathbb{P}_{2,n}(|\wedge(B_2)| = m) = 1/n, \quad 0 \leq m < n.$$

Allgemeiner gilt

**Satz 2.11.** Sei  $\mu = 1, k \geq 2$  und  $(p_j)_{j \geq 0}$  besitze endliches  $k$ -tes Moment. Für die Verteilung der Höhe des MRCA  $|\wedge(B_k)|$  gilt

$$\mathbb{P}_{k,n}(|\wedge(B_k)| \geq m) = \frac{\mathbb{E}[|z_{n-m}(T)|]_k}{\mathbb{E}[|z_n(T)|]_k}, \quad 0 \leq m \leq n. \quad (2.71)$$

**Beweis.** Schritt 1 der Konstruktion liefert eine Darstellung der  $k$ -ten faktoriellen Momente der  $n$ -ten Generationsgröße

$$\mathbb{E}[|z_n(T)|]_k = k! \sum_{\bar{r} \in \mathcal{R}_k^*} c_n(\bar{r}) \prod_{v \in \bar{r}} \frac{f^{(c(v, \bar{r}))}(1)}{c(v, \bar{r})!}. \quad (2.72)$$

Mit der Beobachtung, dass es genau  $c_{n-m}(\bar{r})$  Skelette  $r \in \mathcal{R}_k^n$  gibt, deren MRCA Mindesthöhe  $m$  besitzt und die zu  $\bar{r}$  äquivalent sind, folgern wir

$$\begin{aligned}
\mathbb{P}_{k,n}(| \wedge (B_k) | \geq m) &= \mathbb{P}_n(| \wedge (B_k^*) | \geq m) \\
&= \sum_{\bar{r} \in \mathcal{R}_k^*} \mathbb{P}_n(\bar{R}_k = \bar{r}) \mathbb{P}_n(| \wedge (B_k^*) | \geq m \mid \bar{R}_k = \bar{r}) \\
&= \sum_{\bar{r} \in \mathcal{R}_k^*} \frac{k! c_n(\bar{r}) \prod_{v \in \bar{r}} \frac{f^{(c(v, \bar{r}))}(1)}{c(v, \bar{r})!}}{\mathbb{E}[|z_n(T)|]_k} \frac{c_{n-m}(\bar{r})}{c_n(\bar{r})} \\
&= \frac{\mathbb{E}[|z_{n-m}(T)|]_k}{\mathbb{E}[|z_n(T)|]_k}, \quad 0 \leq m \leq n.
\end{aligned}$$

□

**Bemerkung.** Mittels der Erzeugendenfunktion  $f$  der Kinderzahlverteilung  $(p_j)_{j \geq 0}$  lässt sich (2.71) auch darstellen durch

$$\mathbb{P}_{k,n}(| \wedge (B_k) | \geq m) = \frac{f_{(n-m)}^{(k)}(1)}{f_{(n)}^{(k)}(1)}, \quad 0 \leq m \leq n.$$

**Beispiel: Geometrische Kinderzahlverteilung.**

Betrachte die Kinderzahlverteilung  $(p_j)_{j \geq 0}$  mit

$$p_j = 2^{-(j+1)}, \quad j \geq 0.$$

$(p_j)_{j \geq 0}$  besitzt die Erzeugendenfunktion  $f(s) = (2-s)^{-1}$  und es gilt

$$f^{(m)}(1) = m!, \quad m \geq 1.$$

Mit (2.69) und (2.70) erhält man

$$\mathbb{P}_n(\bar{R}_k = \bar{r}) = \frac{k! c_n(\bar{r})}{\mathbb{E}[|z_n(T)|]_k}, \quad \bar{r} \in \mathcal{R}_k^*$$

und  $R_k^n$  ist uniform auf  $\mathcal{R}_k^n$  verteilt, d.h.

$$\mathbb{P}_n(R_k^n = r) = \frac{k!}{\mathbb{E}[|z_n(T)|]_k} = \frac{1}{n^{k-1}}, \quad r \in \mathcal{R}_k^n.$$

Letztere Gleichheit lässt sich analytisch durch

$$f_{(n)}(s) = \frac{n - (n-1)s}{n+1 - ns}, \quad f_{(n)}^{(k)}(s) = \frac{k! n^{k-1}}{(n+1 - ns)^{k+1}}$$

einsehen oder durch folgende kombinatorische Überlegung: Man erhält einen uniformen geordneten Baum mit genau  $k$  Blättern der Höhe  $n$ , wenn man den

Baum bezüglich der  $k$  Blätter von links nach rechts konstruiert. Man beginnt mit der linkestmöglichen Ahnenlinie die zum ersten Blatt führt und zieht die Generation des MRCA des  $j$ -ten und  $(j-1)$ -ten Blattes ( $2 \leq j \leq k$ ) uniform auf der Ahnenlinie des  $(j-1)$ -ten Blattes. Für jedes der  $k-1$  Blätter rechts des ersten Blattes gibt es somit  $n$  Möglichkeiten, die Generation des Verzweigungsknotens zu wählen.

Weiterhin erhält man mit (2.71) für die Höhe des MRCA

$$\mathbb{P}_{k,n}(|\wedge(B_k)| \geq m) = \left(1 - \frac{m}{n}\right)^{k-1}, \quad 0 \leq m \leq n.$$

Allgemeiner gilt im Limes  $n \rightarrow \infty$ : Die Verteilung der reskalierten Höhe des MRCA konvergiert gegen eine Beta-Verteilung zum Parameter  $(1, k-1)$ . Der Limes des reduzierten  $k$ -Skeletts  $\bar{R}_k$  ist zudem ein binäres Skelett.

**Satz 2.12.** *Sei  $\mu = 1, k \geq 2$  und  $(p_j)_{j \geq 0}$  besitze endliches  $k$ -tes Moment. Dann gilt*

$$\lim_{n \rightarrow \infty} \mathbb{P}_{k,n} \left( \frac{|\wedge(B_k)|}{n} \geq u \right) = (1-u)^{k-1}, \quad 0 \leq u \leq 1 \quad (2.73)$$

und für jedes reduzierte  $k$ -Skelett  $\bar{r} \in \mathcal{R}_k^*$

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(\bar{R}_k = \bar{r}) = \begin{cases} C_{k-1}^{-1}, & \text{für } \beta_2(\bar{r}) = k-1, \\ 0, & \text{sonst.} \end{cases} \quad (2.74)$$

$C_k$  ist die  $k$ -te Catalan-Zahl.

**Beweis.** Wir bestimmen die asymptotische Anzahl binärer Skelette mit  $k$  Blättern in  $n$ -ter Generation für  $n \rightarrow \infty$ . Zunächst gilt  $|\mathcal{R}_k^n| = n^{k-1}$ . Folgende Konstruktion zeigt, dass es mindestens  $[n]_k$  geordnete binäre Bäume in  $\mathcal{R}_k^n$  gibt: Wir beginnen mit der linkestmöglichen Ahnenlinie der Höhe  $n$  und wählen darauf den MRCA zum zweiten Blatt. Den MRCA des  $j$ -ten und  $(j-1)$ -ten Blattes ( $3 \leq j \leq k$ ) bestimmt man nun auf der  $(j-1)$ -ten Ahnenlinie so, dass keiner der maximal  $j-2$  vorhergehend auftretenden Verzweigungsknoten getroffen wird. Mit dieser Konstruktion sieht man zum einen, dass es mindestens  $[n]_k$  geordnete binäre Bäume in  $\mathcal{R}_k^n$  gibt und zum anderen, dass für zwei binäre reduzierte  $k$ -Skelette  $\bar{r}_1$  und  $\bar{r}_2$  aus  $\mathcal{R}_k^*$  gilt  $c_n(\bar{r}_1) \sim c_n(\bar{r}_2)$ ,  $n \rightarrow \infty$ . Wir erhalten

$$\sum_{\bar{r} \in \mathcal{R}_k^* : \bar{r} \text{ binär}} c_n(\bar{r}) \sim n^{k-1}, \quad n \rightarrow \infty$$



und mit (2.72) gilt

$$\mathbb{E}[|z_n(T)|]_k \sim k! n^{k-1} \left( \frac{\sigma^2}{2} \right)^{k-1}, \quad n \rightarrow \infty.$$

Dies liefert für  $0 \leq u \leq 1$

$$\lim_{n \rightarrow \infty} \mathbb{P}_{k,n}(| \wedge (B_k) | \geq nu) = \lim_{n \rightarrow \infty} \frac{\mathbb{E}[|z_{n-nu}(T)|]_k}{\mathbb{E}[|z_n(T)|]_k} = (1-u)^{k-1}.$$

Da für binäre Skelette  $\bar{r}_1$  und  $\bar{r}_2$  aus  $\mathcal{R}_k^*$  gilt  $c_n(\bar{r}_1) \sim c_n(\bar{r}_2)$ ,  $n \rightarrow \infty$ , folgt

$$\lim_{n \rightarrow \infty} \mathbb{P}_n(\bar{R}_k = \bar{r}_1) = \lim_{n \rightarrow \infty} \mathbb{P}_n(\bar{R}_k = \bar{r}_2)$$

und somit (2.74).

□

**Bemerkung.** Die Konstruktion des in der Größe der  $n$ -ten Generation größenverzerrten Baumes, lässt sich auch für eine Kinderzahlverteilung mit unendlicher Varianz betrachten. Man benutzt dazu (siehe Kapitel 2.2) die mit dem Parameter  $\theta$  gekippte Kinderzahlverteilung  $(p_j(\theta))_{j \geq 0}$ .



## Kapitel 3

# Galton-Watson Bäume bedingt auf Mindesthöhe

Wir untersuchen nun die Struktur eines auf Mindesthöhe  $n$  bedingten GW-Baumes. Durch das Bedingen geht zunächst die lokale Unabhängigkeit der Verzweigung verloren. In dem auf Mindesthöhe bedingte Baum kann man aber eine zum Kontext von Kapitel 2 verwandte, Unabhängigkeitsstruktur wiedererlangen, wenn man entlang einer ausgezeichneten Ahnenlinie zerlegt. Wir nutzen die in [15] vorgestellte Konstruktion des bedingten Baumes um klassische Sätze über die Höhe des MRCA sowie die Grenzverteilung der reskalierten  $n$ -ten Generationsgröße für den Fall einer Kinderzahlverteilung mit unendlicher Varianz auf einem alternativen und aufschlussreichen Weg zu beweisen. Dabei erhalten wir zudem die Grenzverteilung der Anzahl der Kinder des MRCA mit Nachkommen in der  $n$ -ten Generation.

Im Falle einer kritischen oder superkritischen Kinderzahlverteilung ( $\mu \geq 1$ ) ist der auf Mindesthöhe bedingte Baum mit dem auf Gesamtgröße bedingten Baum eng verwandt. Asymptotisch stimmen nämlich die bedingten Verteilungen der Stammbäume überein. Für subkritische Verteilungen sind diese jedoch essentiell verschieden (vgl. [4]).

### 3.1 Konstruktion des bedingten GW-Baumes

Zum Verständnis der Konstruktion wird sich folgende einfache Beobachtung als nützlich erweisen: Wir betrachten zunächst einen Baum  $T$  auf dem Ereignis  $\{Z_n > 0\}$ , wobei  $Z_n := |z_n(T)|$  die Anzahl der Knoten in der  $n$ -ten Generation ist. Den linken Knoten der  $n$ -ten Generation erhalten und bezeichnen wir mit  $(V_1^n, \dots, V_n^n) := \min z_n(T)$  (das Minimum einer Knotenmenge ist bezüglich der lexikographischen Ordnung  $\prec$  aufzufassen). Diese ausgezeichnete Ahnenlinie zur  $n$ -ten Generation hat die Eigenschaft, dass alle Knoten in Teilbäumen links der Ahnenlinie die Generation  $n - 1$  nicht überschreiten. Die Ahnenlinie  $(V_1^n, \dots, V_n^n)$  lässt sich bedingt auf dem Ereignis  $\{Z_n > 0\}$  folgendermaßen beschreiben

$$\begin{aligned} V_1^n &= \min\{1 \leq i \leq c(\Delta, T) : z_{n-1}(T_{(i)}) > 0\} \\ V_2^n &= \min\{1 \leq i \leq c((V_1^n), T) : z_{n-2}(T_{(V_1^n, i)}) > 0\} \\ &\vdots \\ V_{n-1}^n &= \min\{1 \leq i \leq c((V_1^n, \dots, V_{n-2}^n), T) : z_1(T_{(V_1^n, \dots, V_{n-2}^n, i)}) > 0\} \\ V_n^n &= 1. \end{aligned}$$

$V_j^n$  ist gerade das jüngste Kind von  $V_{j-1}^n$ , an dem ein Teilbaum der Mindesthöhe  $n - j$  hängt. Das nächste Lemma liefert nun die Grundidee zur Konstruktion des bedingten Baumes. Es beschreibt die bedingte Verteilung der Teilbäume, die aus den Knoten der ersten Generation wachsen.

**Lemma 3.1.** *Bedingt auf das Ereignis  $\{V_1^n = j, Z_1 = k\}$ ,  $1 \leq j \leq k < \infty$ , sind die Teilbäume  $T_{(i)}$ ,  $1 \leq i \leq Z_1$ , stochastisch unabhängig. Die bedingte Verteilung der  $T_{(i)}$  ist*

$$\mathcal{L}(T_{(i)} | V_1^n = j, Z_1 = k) = \begin{cases} \mathcal{L}(T | Z_{n-1} = 0), & 1 \leq i \leq j - 1 \\ \mathcal{L}(T | Z_{n-1} > 0), & i = j \\ \mathcal{L}(T), & j + 1 \leq i \leq k. \end{cases} \quad (3.1)$$

Die bedingte gemeinsame Verteilung von  $(V_1^n, Z_1)$  ist

$$\mathbb{P}(V_1^n = j, Z_1 = k | Z_n > 0) = c_n p_k \mathbb{P}(Z_{n-1} = 0)^{j-1}, \quad (3.2)$$

wobei

$$c_n := \mathbb{P}(Z_{n-1} > 0) / \mathbb{P}(Z_n > 0).$$

Man erkennt nun zum einen die gewonnene Unabhängigkeit und zum anderen eine rekursive Struktur des bedingten Baumes. Bedingt auf die Lage zum ausgezeichneten Knoten, ist die Verzweigung der Individuen in der ersten Generation unabhängig voneinander. Der auf Höhe  $n$  bedingte Baum enthält einen auf Höhe  $n - 1$  bedingten Baum und eine Anzahl bedingter und unbedingter Bäume, die links und rechts der Ahnenlinie herauswachsen. Eine Konstruktion des bedingten Baumes erhält man, wenn man in der Abfolge der Generationen rückwärts geht.

**Konstruktion 3.2.** Folge zufälliger Stammbäume  $\tilde{T}_n$ :

$\tilde{T}_0$ :  $\tilde{T}_0$  sei ein unbedingter GW-Baum mit Kinderzahlverteilung  $(p_k)_{k \geq 0}$ .

$\tilde{T}_1$ : Setze  $\tilde{V}_1^1 := 1$  und ziehe  $\tilde{Y}_1$  gemäß der Verteilung  $\mathbb{P}(\tilde{Y}_1 = k) = p_k(1 - p_0)^{-1}, k \geq 1$ . Füge eine neue Wurzel mit  $\tilde{Y}_1$  Kindern ein und ersetze die Wurzel des Baumes  $\tilde{T}_0$  durch das erste Kind der neuen Wurzel. Füge an die restlichen  $\tilde{Y}_1 - 1$  Kinder unabhängige, kritische GW-Bäume an. Bezeichne den so erhaltenen Baum  $\tilde{T}_1$ .

$\tilde{T}_{n+1}$ : Gegeben  $\tilde{T}_n$ , setze die ausgezeichnete Ahnenlinie  $\tilde{V}_1^n, \dots, \tilde{V}_n^n$  in  $\tilde{T}_n$  am Anfang fort. Definiere dazu  $\tilde{V}_2^{n+1} := \tilde{V}_1^n, \dots, \tilde{V}_{n+1}^{n+1} := \tilde{V}_n^n$  und ziehe  $(\tilde{V}_1^{n+1}, \tilde{Y}_{n+1})$  gemäß

$$\mathbb{P}(\tilde{V}_1^{n+1} = j, \tilde{Y}_{n+1} = k) := c_{n+1} p_k \mathbb{P}(Z_n = 0)^{j-1} \quad (3.3)$$

mit

$$c_{n+1} := \mathbb{P}(Z_n > 0) / \mathbb{P}(Z_{n+1} > 0).$$

Füge eine neue Wurzel mit  $\tilde{Y}_{n+1}$  Kindern ein. Ersetze dabei die Wurzel des Baumes  $\tilde{T}_n$  durch das  $\tilde{V}_1^{n+1}$ -te Kind der neuen Wurzel. Füge an jedes Kind der neuen Wurzel, welches sich links von  $\tilde{V}_1^{n+1}$  befindet, unabhängige kritische GW-Bäume bedingt auf maximale Höhe  $n-1$  ein. An jedes Kind rechts des  $\tilde{V}_1^{n+1}$ -ten Knotens hänge unbedingte, unabhängige kritische GW-Bäume. Bezeichne den so erhaltenen Baum  $\tilde{T}_{n+1}$ .

Die rekursive Konstruktion liefert offenbar eine Folge zufälliger Stammbäume  $(\tilde{T}_n)_{n \geq 0}$  mit der Verteilung

$$\mathcal{L}(\tilde{T}_n) = \mathcal{L}(T \mid Z_n > 0), \quad n \geq 0. \quad (3.4)$$

Den großen Vorteil dieser Konstruktion sieht man, wenn man sich die Folge der Generationsgrößen  $(\tilde{Z}_n)_{n \geq 0}$ , mit  $\tilde{Z}_n := |z_n(\tilde{T}_n)|$ , betrachtet. Zunächst ist  $\mathcal{L}(\tilde{Z}_n) = \mathcal{L}(Z_n | Z_n > 0)$ ,  $n \geq 0$ . Weiterhin sieht man, dass die Knoten der  $n$ -ten Generation von  $\tilde{T}_n$  gerade die Nachkommen in Generation  $n+1$  des ausgezeichneten Knotens erster Generation in  $\tilde{T}_{n+1}$  sind. Alle weiteren Knoten der  $(n+1)$ -ten Generation stammen von Geschwistern zur Rechten des ausgezeichneten Knotens erster Generation. Letzterer Beitrag zur  $(n+1)$ -ten Generation ist nach Lemma 3.1 unabhängig von  $\tilde{T}_n$ . Mit  $\tilde{M}_n$  bezeichnen wir die Anzahl der älteren Geschwister des ausgezeichneten Knotens erster Generation in der Konstruktion von  $\tilde{T}_n$ , d.h. die Zufallsvariablen  $\tilde{M}_n$ ,  $n \geq 1$  sind unabhängig mit Verteilung

$$\mathcal{L}(\tilde{M}_n) = \mathcal{L}(\tilde{Y}_n - \tilde{V}_1^n),$$

d.h. mit (3.3) gilt

$$\mathbb{P}(\tilde{M}_n = k) = c_n \sum_{j \geq k+1} p_j \mathbb{P}(Z_{n-1} = 0)^{j-k-1}, \quad k \geq 0. \quad (3.5)$$

Damit folgt, dass  $(\tilde{Z}_n)_{n \geq 0}$  unabhängige, nichtnegative Zuwächse besitzt mit Verteilung

$$\mathcal{L}(\tilde{Z}_{n+1} - \tilde{Z}_n) = \mathcal{L}\left(\sum_{i=1}^{\tilde{M}_{n+1}} Z_{n,i}\right), \quad (3.6)$$

wobei  $Z_{n,i}$ ,  $i \geq 1$ ,  $n \geq 0$  unabhängige Zufallsvariablen mit Verteilung  $\mathcal{L}(Z_{n,i}) = \mathcal{L}(Z_n)$  und unabhängig von  $\tilde{M}_n$ ,  $n \geq 1$  sind. Gleichung (3.6) wird grundlegend bei der weiteren Analyse des bedingten Baumes sein.

Offenbar kann man die Konstruktion des Baumes  $\tilde{T}_n$  unabhängig der Kinderzahlverteilung  $(p_k)_{k \geq 0}$  durchführen. Man erhält aber mit dieser Konstruktion für kritische und subkritische Kinderzahlverteilungen aufschlussreichere Erkenntnisse, da hier der Baum nach dem Kritizitätstheorem 1.1 fast sicher endlich ist.

Im weiteren Verlauf setzen wir eine kritische Verteilung voraus.

Wir betrachten nun die Asymptotik der ausgezeichneten Ahnenlinie für  $n \rightarrow \infty$ .

Für die Zahlen  $c_n$  in (3.2) gilt

$$\begin{aligned} c_n &= \frac{\mathbb{P}(Z_{n-1} > 0)}{\mathbb{P}(Z_n > 0)} = \frac{1 - \mathbb{P}(Z_{n-1} = 0)}{1 - f(\mathbb{P}(Z_{n-1} = 0))} \\ &\rightarrow f'(1)^{-1} = 1, \quad n \rightarrow \infty. \end{aligned} \quad (3.7)$$

Im Limes  $n \rightarrow \infty$  erhält man unter Berücksichtigung von  $\mathbb{P}(Z_n > 0) \rightarrow 0$

$$\lim_{n \rightarrow \infty} \mathbb{P}(V_1^n = j, Z_1 = k | Z_n > 0) = p_k$$

und somit

$$\lim_{n \rightarrow \infty} \mathbb{P}(Z_1 = k | Z_n > 0) = kp_k, \quad k \geq 1, \quad (3.8)$$

$$\lim_{n \rightarrow \infty} \mathbb{P}(V_1^n = j | Z_1 = k, Z_n > 0) = 1/k, \quad 1 \leq j \leq k. \quad (3.9)$$

Asymptotisch ist die Verteilung der Kinderzahlen der ausgezeichneten Ahnenlinie also einfach größenverzerrt. Die Position des ausgezeichneten Knotens der Ahnenlinie unter den Geschwistern ist auf Familiengröße bedingt uniform verteilt.

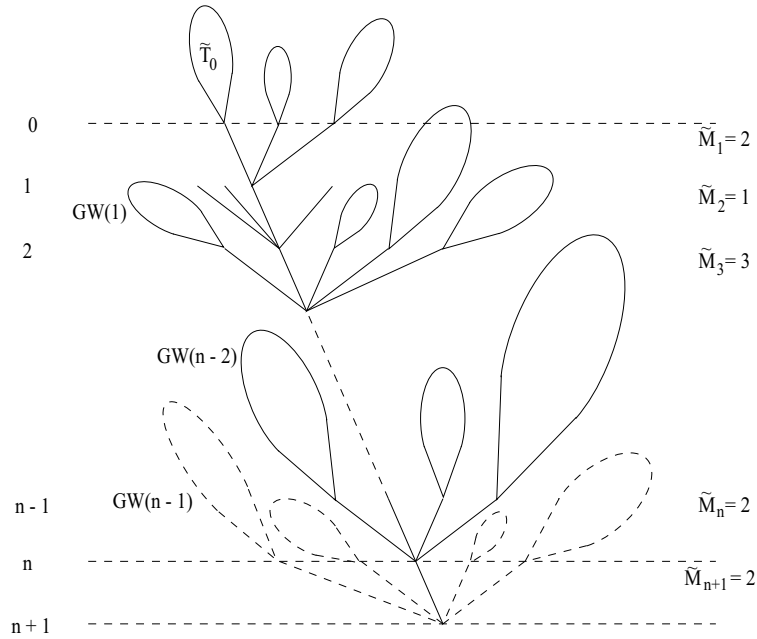


ABBILDUNG 4.1: Konstruktion des Baumes  $\tilde{T}_{n+1}$  aus  $\tilde{T}_n$ . Die Teilbäume  $GW(k)$  sind auf maximale Höhe  $k$  bedingte GW-Bäume. Die Teilbäume rechts der ausgezeichneten Ahnenlinie sowie der Baum  $\tilde{T}_0$  sind unbedingte GW-Bäume.

### 3.2 Grenzwertsätze über den bedingten GW-Baum

Wir interessieren uns nun für den jüngsten gemeinsamen Vorfahren aller Knoten  $n$ -ter Generation (zur Notation vgl. (1.9))

$$A_n := \wedge(z_n(T))$$

und dessen Kinder mit Nachfolgern in der  $n$ -ten Generation

$$K_n := \{v \in T : |v| = |A_n| + 1, v \wedge z_n(T) = v\}.$$

Dazu untersuchen wir die bedingten Verteilungen

$$\mathcal{L}(|A_n|/n \mid Z_n > 0) \quad \text{und} \quad \mathcal{L}(|K_n| \mid Z_n > 0)$$

im Limes  $n \rightarrow \infty$ . Zudem betrachten wir die Grenzverteilung der geeignet skalierten, bedingten  $n$ -ten Generation

$$\lim_{n \rightarrow \infty} \mathcal{L}(Z_n/\rho_n \mid Z_n > 0).$$

Wir zitieren nun die klassischen Grenzwertsätze und beweisen diese im Licht der eben vorgestellten Konstruktion. Dabei nutzen wir ähnliche Beweisideen wie in [15] unter Zuhilfenahme zusätzlicher analytischer Hilfsmittel.

#### Generation und Kinderzahl des MRCA

Wir beginnen mit einem Resultat von Zubkov [33] das besagt, dass im kritischen, auf Mindesthöhe  $n$  bedingten GW-Baum die Generation des MRCA aller Knoten  $n$ -ter Generation asymptotisch uniform auf der Mindesthöhe verteilt ist. Dies beweisen wir für Kinderzahlverteilungen endlicher Varianz, sowie für solche, die im Anziehungsbereich einer stabilen Verteilung sind. Offen bleibt, ob allein schon die Kritizität der Verteilung ausreicht.

Die hier zitierten klassischen Sätze setzen eine kritische Kinderzahlverteilung mit einer Erzeugendenfunktion der Gestalt

$$f(s) = s + (1-s)^\alpha L(1/(1-s)), \quad 0 \leq s \leq 1 \quad (3.10)$$

voraus, wobei  $L$  eine langsam variierende Funktion ist und  $1 < \alpha \leq 2$ . Wir werden am Ende dieses Abschnitts zeigen, dass  $(p_j)_{j \geq 0}$  genau dann eine Erzeugendenfunktion der Gestalt (3.10) besitzt, wenn sie im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung liegt. Wir formulieren daher die nachfolgenden Aussagen unter dieser äquivalenten Voraussetzung.



**Satz 3.3.** (Zubkov, 1975) Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|A_n|/n \leq u \mid Z_n > 0) = u, \quad 0 \leq u \leq 1. \quad (3.11)$$

Für die Verteilung der Kinder des MRCA mit Nachkommen in  $n$ -ter Generation zeigt sich, dass im Fall endlicher Varianz nur zwei Kinder zum Limes  $n \rightarrow \infty$  beitragen: der Nachfolger des MRCA auf der ausgezeichneten Ahnenlinie und ein weiteres Geschwister. Hingegen erhält man im Fall unendlicher Varianz die Grenzverteilung (2.28) bedingt auf mindestens 2 Kinder. Dieses Resultat findet man unter anderem Zugang in [10] (Theorem 2.7.1).

**Theorem 3.4.** Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P}(|K_n| = k \mid Z_n > 0) = q_k(\alpha)(1 - 1/\alpha)^{-1}, \quad k \geq 2, \quad (3.12)$$

mit

$$q_k(\alpha) := \frac{\alpha - 1}{\Gamma(2 - \alpha)} \frac{\Gamma(k - \alpha)}{\Gamma(k + 1)}, \quad k \geq 2, \quad (3.13)$$

wobei

$$\Gamma(0)/\Gamma(0) := 1 \quad \text{und} \quad \Gamma(j)/\Gamma(0) := 0 \quad \text{für} \quad j \geq 1.$$

**Bemerkung.** Die Zerlegung der  $n$ -ten Generation nach den Beiträgen der Kinder des MRCA wird im nächsten Kapitel der Zugang zu einem alternativen Beweis über die Grenzverteilung der  $n$ -ten Generation sein.

Für die Beweise benötigen wir die Asymptotik der Überlebenswahrscheinlichkeiten eines kritischen GW-Prozesses. Die folgende Version der Kolmogorovschen Asymptotik ohne Voraussetzungen an höhere Momente wurde in [23] bewiesen. Einen auf Konstruktion (3.3) beruhenden eleganten und elementaren Beweis findet man in [15].

**Satz 3.5.** (Kesten, Ney und Spitzer, 1966) Sei  $\mu = 1$  und  $0 < \sigma^2 < \infty$ , dann gilt

$$\mathbb{P}(Z_n > 0) \sim \frac{2}{\sigma^2 n} \quad \text{für} \quad n \rightarrow \infty. \quad (3.14)$$

Eine Asymptotik für den Fall eines stabilen Anziehungsbereiches liefert das klassische Lemma von Slack [30].

**Lemma 3.6.** (Slack, 1967) Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt

$$\mathbb{P}(Z_n > 0)^{\alpha-1} L(\mathbb{P}(Z_n > 0)^{-1}) \sim \frac{1}{(\alpha-1)n} \text{ für } n \rightarrow \infty, \quad (3.15)$$

wobei  $L$  die langsam variierende Funktion aus Darstellung (3.10) ist.

**Bemerkung.** Für den Fall  $\alpha = 2$  mit  $\sigma^2 < \infty$  erhält man mit  $L(s) \sim \sigma^2/2$ ,  $s \rightarrow \infty$  die Asymptotik von Kolmogorov.

**Beweis von Satz 3.3.** Sei  $\tilde{A}_n$  der MRCA aller Knoten der  $n$ -ten Generation in  $\tilde{T}_n$ . Aufgrund (3.4) ist  $\mathcal{L}(\tilde{A}_n) = \mathcal{L}(A_n \mid Z_n > 0)$ . Nach Konstruktion (3.2) befindet sich  $\tilde{A}_n$  auf der ausgezeichneten Ahnenlinie offenbar genau dann in Generation  $n - k$ , falls  $\tilde{Z}_n = \tilde{Z}_k$ , d.h. falls es in der Konstruktion von  $\tilde{T}_n$  nach der Konstruktion von  $\tilde{T}_k$  keinen weiteren Zuwachs mehr zur obersten Generation gibt. Aus dieser Beobachtung und der Unabhängigkeitsstruktur (3.6) folgt

$$\mathbb{P}(|\tilde{A}_n| \geq n - k) = \prod_{j=k}^{n-1} \mathbb{P}(\tilde{Z}_{j+1} = \tilde{Z}_j). \quad (3.16)$$

Für die Faktoren in (3.16) gilt mit (3.5)

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{j+1} = \tilde{Z}_j) &= \mathbb{P}(Z_{j,1} + \dots + Z_{j,\tilde{M}_{j+1}} = 0) \\ &= \sum_{r=0}^{\infty} \mathbb{P}(\tilde{M}_{j+1} = r) \mathbb{P}(Z_j = 0)^r \\ &= c_{j+1} \sum_{r=0}^{\infty} \sum_{l=r+1}^{\infty} p_l \mathbb{P}(Z_j = 0)^{l-(r+1)} \mathbb{P}(Z_j = 0)^r \\ &= c_{j+1} \sum_{l=1}^{\infty} l p_l \mathbb{P}(Z_j = 0)^{l-1} \\ &= c_{j+1} f'(\mathbb{P}(Z_j = 0)). \end{aligned} \quad (3.17)$$

Wir zeigen nun

$$\lim_{j \rightarrow \infty} j \mathbb{P}(\tilde{Z}_{j+1} > \tilde{Z}_j) = 1. \quad (3.18)$$

Aus (3.18) und (3.16) folgt damit

$$\lim_{n \rightarrow \infty} \mathbb{P}(|\tilde{A}_n| \geq n - k_n) \sim k_n/n, \quad 0 \leq k_n \leq n, \quad k_n, n \uparrow \infty \quad (3.19)$$

und somit Behauptung (3.11).

Zunächst gilt nach (3.3), (3.17) und mit  $\rho_{j+1} := \mathbb{P}(Z_{j+1} > 0) = 1 - f(1 - \rho_j)$

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{j+1} > \tilde{Z}_j) &= 1 - c_{j+1} f'(\mathbb{P}(Z_j = 0)) \\ &= \frac{1 - f(1 - \rho_j) - \rho_j f'(1 - \rho_j)}{\rho_{j+1}}. \end{aligned}$$

Im Fall  $\sigma^2 < \infty$  liefert eine Taylorentwicklung

$$\begin{aligned} 1 - f(1 - \rho_j) - \rho_j f'(1 - \rho_j) &= 1 - (f(1) - f'(1)\rho_j + f''(1)\rho_j^2 + O(\rho_j^3)) - \rho_j(f'(1) - f''(1)\rho_j + O(\rho_j^2)) \\ &= \frac{\sigma^2}{2}\rho_j^2 + O(\rho_j^3). \end{aligned}$$

Mit Hilfe der Kolmogorovschen Asymptotik (3.14) erhält man Behauptung (3.18)

$$\begin{aligned} \mathbb{P}(\tilde{Z}_{j+1} > \tilde{Z}_j) &\sim \frac{\sigma^2}{2} \rho_j \frac{\rho_j}{\rho_{j+1}} \\ &\sim \frac{1}{j}, \quad j \rightarrow \infty. \end{aligned}$$

Im Fall unendlicher Varianz gilt nach Satz 3.7, (3.25) und (3.26)

$$\begin{aligned} 1 - f(1 - \rho_j) - \rho_j f'(1 - \rho_j) &= \rho_j - \rho_j^\alpha L(1/\rho_j) - \rho_j f'(1 - \rho_j) \\ &= -\rho_j^\alpha L(1/\rho_j) + \rho_j(1 - f'(1 - \rho_j)) \\ &\sim (\alpha - 1) \rho_j^\alpha L(1/\rho_j), \quad j \rightarrow \infty. \end{aligned}$$

Die Slacksche Asymptotik (3.15) liefert schließlich

$$\mathbb{P}(\tilde{Z}_{j+1} > \tilde{Z}_j) \sim \frac{\rho_j}{\rho_{j+1}} (\alpha - 1) \rho_j^{\alpha-1} L(1/\rho_j) \sim \frac{1}{j}, \quad j \rightarrow \infty.$$

□

**Beweis von Theorem 3.4.** Aufgrund (3.4) genügt es, die Aussage für die Kinder des MRCA im Baum  $\tilde{T}_n$  zu beweisen.  $\tilde{A}_n$  sei der MRCA aller Knoten der  $n$ -ten Generation von  $\tilde{T}_n$  und  $\tilde{K}_n$  die Anzahl der Kinder von  $\tilde{A}_n$  mit Nachkommen in Generation  $n$ . Wir betrachten nun die gemeinsame Verteilung von  $|\tilde{A}_n|$  und  $\tilde{K}_n$ . Nach Konstruktion befindet sich  $\tilde{A}_n$  in Generation  $n - (k + 1)$  und  $\tilde{K}_n = r + 1$  genau dann, wenn  $\tilde{Z}_n = \tilde{Z}_{k+1}$  und wenn an genau  $r$  der  $\tilde{M}_{k+1}$  Geschwister zur Rechten des ausgezeichneten Kindes von  $\tilde{A}_n$  Teilbäume mit

Mindesthöhe  $k$  hängen. Mit dem Satz von der totalen Wahrscheinlichkeit gilt zunächst

$$\mathbb{P}(\tilde{K}_n = r + 1) = \sum_{k=0}^{n-1} \mathbb{P}(\tilde{K}_n = r + 1, |\tilde{A}_n| = n - (k + 1)). \quad (3.20)$$

Mittels (3.5) erhält man

$$\begin{aligned} & \mathbb{P}(\tilde{K}_n = r + 1, |\tilde{A}_n| = n - (k + 1)) \\ &= \mathbb{P}(\tilde{Z}_n = \tilde{Z}_{k+1}) \sum_{m=r}^{\infty} \binom{m}{r} \mathbb{P}(\tilde{M}_{k+1} = m) \mathbb{P}(Z_k = 0)^{m-r} \mathbb{P}(Z_k > 0)^r \\ &= P(\tilde{Z}_n = \tilde{Z}_{k+1}) c_{k+1} \sum_{m=r}^{\infty} \binom{m}{r} \sum_{j=m+1}^{\infty} p_j (1 - \rho_k)^{j-m-1} (1 - \rho_k)^{m-r} \rho_k^r \\ &= P(\tilde{Z}_n = \tilde{Z}_{k+1}) \frac{\rho_k^{r+1}}{\rho_{k+1}} \sum_{j=r+1}^{\infty} p_j (1 - \rho_k)^{j-(r+1)} \sum_{m=r}^{j-1} \binom{m}{r} \\ &= P(\tilde{Z}_n = \tilde{Z}_{k+1}) \frac{\rho_k^{r+1}}{(r+1)! \rho_{k+1}} \sum_{j=r+1}^{\infty} [j]_{r+1} p_j (1 - \rho_k)^{j-(r+1)} \\ &= \mathbb{P}(\tilde{Z}_n = \tilde{Z}_{k+1}) \frac{\rho_k^{r+1} f^{(r+1)}(1 - \rho_k)}{(r+1)! \rho_{k+1}}. \end{aligned} \quad (3.21)$$

Weiterhin folgt mit (3.26), (2.31) und (3.15) für den Quotienten in (3.21) im Fall  $1 < \alpha < 2$

$$\begin{aligned} & \frac{\rho_k^{r+1} f^{(r+1)}(1 - \rho_k)}{(r+1)! \rho_{k+1}} \\ & \sim \frac{\rho_k^{r+1}}{(r+1)! \rho_{k+1}} \frac{\alpha(\alpha-1)\Gamma(r+1-\alpha)}{\Gamma(2-\alpha)} \rho_k^{-(r+1-\alpha)} L(1/\rho_k) \\ & \sim \alpha q_{r+1}(\alpha) \rho_k^{\alpha-1} L(1/\rho_k) \\ & \sim \frac{\alpha}{\alpha-1} q_{r+1}(\alpha) \frac{1}{k}, \quad k \rightarrow \infty. \end{aligned} \quad (3.22)$$

Im Fall  $\alpha = 2$ ,  $\sigma^2 = \infty$  folgt mit (2.36), (2.39) und (3.15)

$$\begin{aligned} \frac{\rho_k^{r+1} f^{(r+1)}(1 - \rho_k)}{(r+1)! \rho_{k+1}} & \sim \rho_k L(1/\rho_k) \frac{\rho_k^{r-1} f^{(r+1)}(1 - \rho_k)}{(r+1)! L(1/\rho_k)}, \quad k \rightarrow \infty \\ & = o(1/k), \quad r \geq 2. \end{aligned} \quad (3.23)$$

Mit Satz 3.3 erhält man aus (3.22) bzw. (3.23) und (3.21)

$$\begin{aligned} & \mathbb{P}(\tilde{K}_n = r + 1, |\tilde{A}_n| = n - (k + 1)) \\ & \sim q_{r+1}(\alpha) (1 - 1/\alpha)^{-1} \frac{1}{n}, \quad k < n, \quad k, n \rightarrow \infty. \end{aligned} \quad (3.24)$$

Aus letzterer Asymptotik erkennt man, dass die Generation des MRCA  $|\tilde{A}_n|$  und die Kinderzahl  $\tilde{K}_n$  asymptotisch unabhängig sind. Mit (3.20) folgt schließlich die Behauptung.  $\square$

**Lemma 3.7.** *Eine kritische Kinderzahlverteilung  $(p_k)_{k \geq 0}$  liegt genau dann im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung mit  $1 < \alpha \leq 2$ , wenn es eine langsam variierende Funktion  $L$  gibt, so dass die Erzeugendenfunktion der Kinderzahlverteilung die Gestalt (3.10) besitzt.*

**Beweis.** Sei  $(p_k)_{k \geq 0}$  im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung. Definiere

$$L(1/(1-s)) := \frac{f(s) - s}{(1-s)^\alpha}.$$

Zu zeigen ist, dass  $L$  langsam variiert. Im Fall  $\alpha = 2$  und  $\sigma^2 < \infty$  sieht man dies unmittelbar ein,

$$\frac{f(s) - s}{(1-s)^2} \sim \frac{1 - f'(s)}{2(1-s)} \sim \frac{f''(s)}{2} \sim \frac{\sigma^2}{2}, \quad s \uparrow 1.$$

Im Fall  $\alpha < 2$  (der Beweis für  $\alpha = 2, \sigma^2 = \infty$  ist analog) existiert nach Lemma 2.5 eine langsam variierende Funktion  $\bar{L}$ , so dass für  $s \uparrow 1$

$$1 - \mu(s) \sim \frac{\Gamma(2-\alpha)}{\alpha-1} (1-s)^{\alpha-1} \bar{L}(1/(1-s)).$$

Aus Definition 2.4 sieht man, dass

$$1 - \mu(s) \sim f(s) - sf'(s).$$

Weiterhin folgt aus

$$\lim_{s \uparrow 1} \frac{1 - f'(s)}{f(s) - sf'(s)} = \lim_{s \uparrow 1} \frac{-f''(s)}{-sf''(s)} = 1$$

die Asymptotik

$$1 - f'(s) \sim \frac{\Gamma(2-\alpha)}{\alpha-1} (1-s)^{\alpha-1} \bar{L}(1/(1-s)). \quad (3.25)$$

Mit Hilfe des Theorems von Karamata erhält man schließlich

$$\begin{aligned} f(s) - s &= \int_s^1 (1 - f'(u)) du \\ &= \int_{(1-s)^{-1}}^\infty (1 - f'(1 - 1/u)) \frac{du}{u^2} \\ &\sim \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} (1-s)^\alpha \bar{L}(1/(1-s)) \end{aligned}$$

und es gilt

$$L(1/(1-s)) \sim \frac{\Gamma(2-\alpha)}{\alpha(\alpha-1)} \bar{L}(1/(1-s)). \quad (3.26)$$

$(p_k)_{k \geq 0}$  besitze nun eine Erzeugendenfunktion der Gestalt (3.10). Da  $f$  und somit auch  $L(1/(1-s))$  unendlich oft in  $(0,1)$  differenzierbar sind, erhält man mit Konstanten  $c_0, c_1, c_2$

$$f''(s) = \sum_{j=0}^2 c_j (1-s)^{\alpha-2-j} L^{(j)}(1/(1-s)).$$

Nach Theorem 1.3.3 in [8] gilt für  $L \in C^\infty([1, \infty))$  und  $h(x) := \log L(e^x)$  für alle  $j \geq 1$

$$h^{(j)}(x) \rightarrow 0, \quad x \rightarrow \infty.$$

Damit sieht man, dass

$$f''(s) \sim \alpha(\alpha-1)(1-s)^{\alpha-2} L(1/(1-s)), \quad s \uparrow 1.$$

Mit einem Taubersatz für Potenzreihen (z.B. Theorem 5 in Kapitel XIII.5, [13]) folgt, dass die gestutzten zweiten Momente regulär zum Index  $2-\alpha$  variieren. Dies ist nach Theorem 1, Kapitel IX.8 in [13] äquivalent dazu, dass  $(p_k)_{k \geq 0}$  im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung liegt.

□

## Grenzverteilung der Generationsgröße

Wir untersuchen nun die Grenzverteilung der auf Überleben bedingten, reskalierten  $n$ -ten Generationsgröße. Als natürliche Skalierung stellt sich hier die bedingte erwartete Generationsgröße  $\mathbb{E}(Z_n | Z_n > 0)$  heraus. Aufgrund der Gleichung ( $\mu = 1$ )

$$1 = \mathbb{E}Z_n = \mathbb{E}(Z_n; Z_n > 0) = \mathbb{E}(Z_n | Z_n > 0) \mathbb{P}(Z_n > 0)$$

ist diese Skalierung gerade  $\rho_n^{-1} = \mathbb{P}(Z_n > 0)^{-1}$ .

Das Exponentialgesetz von Yaglom besagt nun, dass die Grenzverteilung im Fall endlicher Varianz eine Exponentialverteilung ist. Yaglom [32] konnte dies unter Annahme eines endlichen dritten Moments der Kinderzahlverteilung beweisen. Einen Beweis, der nur ein endliches zweites Moment voraussetzt, wurde von Kesten, Ney und Spitzer [23] geliefert.

**Satz 3.8.** (*Kesten, Ney und Spitzer, 1966*) Sei  $\mu = 1$  und  $0 < \sigma^2 < \infty$ , dann gilt

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \frac{Z_n}{n} \geq x \mid Z_n > 0 \right) = \exp \left( -\frac{2}{\sigma^2} x \right), \quad x \geq 0. \quad (3.27)$$

Für den Fall eines stabilen Anziehungsbereiches gilt (siehe [30])

**Theorem 3.9.** (*Slack, 1967*) Sei  $\mu = 1$  und  $(p_j)_{j \geq 0}$  liege im Anziehungsbereich einer stabilen Verteilung mit Index  $1 < \alpha \leq 2$ . Dann gilt

$$\mathcal{L}(\rho_n Z_n \mid Z_n > 0) \xrightarrow{w} \mathcal{L}(Z), \quad \text{für } n \rightarrow \infty \quad (3.28)$$

mit der Laplace-Transformierten

$$\mathbb{E} \exp(-\lambda Z) = 1 - \frac{\lambda}{(1 + \lambda^{\alpha-1})^{1/(\alpha-1)}}. \quad (3.29)$$

**Bemerkung.** Einen Beweis von Satz 3.8 mittels Konstruktion 3.2 findet man in [16]. Wir geben hier einen Beweis von Theorem 3.9, der analoge Beweisideen benutzt.

**Beweis von Theorem 3.9.** Wesentlich für den Beweis ist die Zerlegung der Größe der  $n$ -ten Generation  $\tilde{Z}_n$  in eine Summe von unabhängigen Beiträgen der Kinder des MRCA. Mit Hilfe der sogenannten Kontraktionsmethode [27] lässt sich aus der skalierten, rekursiven Beziehung in einer passenden Metrik die Konvergenz von  $\rho_n \tilde{Z}_n$  gegen eine Zufallsvariable  $\bar{Z}$  zeigen. Deren Verteilung ist eindeutig durch die aus der Zerlegung gewonnenen Fixpunktgleichung charakterisiert. Die Fixpunktgleichung ist nämlich, aufgefasst als Abbildung vom Raum der Wahrscheinlichkeitsmaße auf  $\mathbb{R}^+$  in denselben (mit gewissen Bedingungen an die Momente), eine Kontraktion bzgl. der Metrik. Während im Fall endlicher Varianz die  $L_2$ -Wasserstein (Mallow) Metrik die geeignete Metrik ist (siehe [16]), benötigt man im Fall unendlicher Varianz die sogenannte Zolotarev-Metrik [34]. Konvergenz in der Zolotarev-Metrik impliziert schließlich die schwache Konvergenz der Maße. Aus der Fixpunktgleichung charakterisieren wir dann die Grenzverteilung von  $\bar{Z}$  durch eine einfache Berechnung der Laplace-Transformierten. Wir werden hier nicht näher auf die Zolotarev Metrik eingehen sondern kontrollieren nur die Voraussetzungen zur Anwendung des Kontraktionstheorems in [27].

Aus Konstruktion 3.2 sieht man zunächst, dass die Anzahl der Knoten der  $n$ -ten Generation in Verteilung gleich der Summe unabhängiger Beiträge der Nachkommen des MRCA ist, d.h.

$$\mathcal{L}(\tilde{Z}_n) = \mathcal{L}\left(\sum_{i=1}^{\tilde{K}_n} \tilde{Z}_{n-|\tilde{A}_n|}^{(i)}\right),$$

wobei  $\tilde{Z}_k^{(i)}$  unabhängige Zufallsvariablen, unabhängig von  $\tilde{A}_n$  und  $\tilde{K}_n$ , mit Verteilung  $\mathcal{L}(\tilde{Z}_k^{(i)}) = \mathcal{L}(\tilde{Z}_k)$  sind. Durch Reskalierung mit  $\rho_n^{-1}$  und Erweiterung mit  $\rho_{n-|\tilde{A}_n|}$  erhält man

$$\mathcal{L}(\rho_n \tilde{Z}_n) = \mathcal{L}\left(\frac{\rho_n}{\rho_{n-|\tilde{A}_n|}} \sum_{i=1}^{|\tilde{K}_n|} \rho_{n-|\tilde{A}_n|} \tilde{Z}_{n-|\tilde{A}_n|}^{(i)}\right).$$

Dies lässt sich mit der Normierung  $\bar{Z}_n := \rho_n \tilde{Z}_n$  und  $\bar{R}_n := \rho_n / \rho_{n-|\tilde{A}_n|}$  darstellen durch

$$\mathcal{L}(\bar{Z}_n) = \mathcal{L}\left(\bar{R}_n \sum_{i=1}^{|\tilde{K}_n|} \tilde{Z}_{n-|\tilde{A}_n|}^{(i)}\right). \quad (3.30)$$

Aufgrund Asymptotik (3.15) gibt es eine langsam variierende Funktion  $L^*$  (vgl. Kapitel 8.12.3 in [8]) mit

$$\rho_n = n^{-\frac{1}{\alpha-1}} L^*(n).$$

Somit gilt

$$\bar{R}_n = \left(1 - \frac{|\tilde{A}_n|}{n}\right)^{\frac{1}{\alpha-1}} \frac{L^*(n)}{L^*(n - |\tilde{A}_n|)}. \quad (3.31)$$

Mit der gleichmäßigen Konvergenz langsam variierender Funktionen (siehe z.B. Theorem 1.2.1 in [8]) sieht man, dass der Quotient der langsam variierenden Funktionen in (3.31) bedingt auf dem Ereignis  $\{\tilde{A}_n \leq (1 - \epsilon)n\}$  für beliebiges  $\epsilon$  mit  $0 < \epsilon \leq 1$  stochastisch gegen 1 konvergiert. Mit Satz 3.3 gilt

$$\mathcal{L}(\bar{R}_n) \xrightarrow{w} \mathcal{L}\left(U^{\frac{1}{\alpha-1}}\right) \quad (3.32)$$

mit einer auf  $[0, 1]$  uniform verteilten Zufallsvariable  $U$ . Formal erhalten wir im Limes  $n \rightarrow \infty$  aus (3.30) die Fixpunktgleichung

$$\mathcal{L}(\bar{Z}) = \mathcal{L}\left(U^{\frac{1}{\alpha-1}} \sum_{i=1}^{\bar{K}} \bar{Z}^{(i)}\right), \quad (3.33)$$



wobei  $\bar{K}$  die Verteilung aus Theorem 3.4 besitzt,  $U$  uniform auf  $[0, 1]$  verteilt ist und  $\bar{Z}^{(i)}$  unabhängige Kopien von  $\bar{Z}$  sind. Zudem sind  $\bar{K}, U, \bar{Z}^{(1)}, \bar{Z}^{(2)}, \dots$  unabhängig voneinander.

Theorem 4.6 aus [27] angewandt auf Gleichung (3.30) liefert die schwache Konvergenz von  $\bar{Z}_n$  gegen  $\bar{Z}$ . Die Verteilung von  $\bar{Z}$  ist zudem unter allen Verteilungen mit Erwartungswert 1 und endlichen  $s$ -ten Momenten eindeutig durch Fixpunktgleichung (3.33) charakterisiert. An dieser Stelle seien nur die Voraussetzungen zur Anwendung des Kontraktionsprinzips überprüft. Für ein geeignetes  $0 < s \leq 3$  ist

$$\begin{aligned} \mathbb{E}|\bar{Z}_n|^s &< \infty, \quad n \geq 0, \\ \mathbb{E}\left|\left|\bar{R}_n - U^{1/(\alpha-1)}\right|\max(\tilde{K}_n, \bar{K})\right|^s &\rightarrow 0, \quad n \rightarrow \infty, \\ \mathbb{E}|\bar{K}U^{s/(\alpha-1)}| &< 1, \\ \forall l \in \mathbb{N}: \quad \mathbb{E}|I_{\{|\tilde{A}_n| \geq n-l\} \cup \{|\tilde{A}_n|=0\}} \tilde{K}_n \bar{R}_n|^s &\rightarrow 0, \quad n \rightarrow \infty, \end{aligned}$$

zu zeigen. Sei dazu  $1 < s < \alpha$ . Da  $(p_j)_{j \geq 0}$  im Anziehungsbereich einer  $\alpha$ -stabilen Verteilung liegt, gilt für eine Zufallsvariable  $X$  mit dieser Verteilung, dass  $\mathbb{E}X^s = \mathbb{E}Z_1^s < \infty$  und somit  $\mathbb{E}\bar{Z}_1^s = \mathbb{P}(X > 0)^{-1}\mathbb{E}X^s < \infty$ . Weiterhin erhält man für  $\mathbb{E}\bar{Z}_n^s = \mathbb{E}[Z_n^s | Z_n > 0] = \rho_n^{-1}\mathbb{E}[Z_n^s]$  mit unabhängigen Kopien  $X_1, X_2, \dots$  von  $X$

$$\begin{aligned} \mathbb{E}\bar{Z}_n^s &= \rho_n^{-1}\mathbb{E}\left(\sum_{i=1}^{Z_{n-1}} X_i\right)^s \\ &\leq \rho_n^{-1}\mathbb{E}\left(Z_{n-1}^{s-1} \sum_{i=1}^{Z_{n-1}} X_i^s\right) \\ &= \rho_n^{-1}\mathbb{E}\left(\mathbb{E}\left(Z_{n-1}^{s-1} \sum_{i=1}^{Z_{n-1}} X_i^s \mid Z_{n-1}\right)\right) \\ &= \rho_n^{-1}\mathbb{E}Z_{n-1}^s \mathbb{E}X^s < \infty \end{aligned}$$

Dabei gilt obige Ungleichung aufgrund der Konvexität von  $x \mapsto x^s$ ,  $s > 1$ .

Zum Nachweis der zweiten Bedingung bemerken wir zuerst, dass  $\bar{K}$  im normalen Anziehungsbereich einer  $\alpha$ -stabilen Verteilung liegt und somit alle Momente  $s' < \alpha$  besitzt. Sei  $s < s' < \alpha$ . Es gilt mit (3.32)

$$\begin{aligned}
 & \mathbb{E} \left| |\bar{R}_n - U^{1/(\alpha-1)}| \max(\tilde{K}_n, \bar{K}) \right|^s \\
 & \leq \left( \mathbb{E} |\bar{R}_n - U^{1/(\alpha-1)}|^{s'/s} \right)^{\frac{s'-s}{s}} \left( \left( \mathbb{E} K_n^{s'} \right)^{\frac{s}{s'}} + \left( \mathbb{E} K^{s'} \right)^{\frac{s}{s'}} \right) \\
 & \rightarrow 0, \quad n \rightarrow \infty.
 \end{aligned}$$

Weiterhin ist die Kontraktionsbedingung erfüllt:

$$\mathbb{E} |\bar{K} U^{s/(\alpha-1)}| = \mathbb{E} \bar{K} \mathbb{E} U^{s/(\alpha-1)} = \frac{\alpha}{\alpha-1} \frac{\alpha-1}{s+\alpha-1} = \frac{\alpha}{s+\alpha-1} < 1.$$

Die letzte Bedingung kontrolliert man mittels (3.31) und (3.21). Für jedes  $l \in \mathbb{N}$  gilt

$$\begin{aligned}
 & \mathbb{E} |I_{\{|\tilde{A}_n| \geq n-l\} \cup \{|\tilde{A}_n|=0\}} \tilde{K}_n \bar{R}_n|^s \\
 & \leq (l/n)^{s/(\alpha-1)} \mathbb{E} I_{\{|\tilde{A}_n| \geq n-l\}} \tilde{K}_n^s + \mathbb{E} I_{\{|\tilde{A}_n|=0\}} \tilde{K}_n^s \\
 & \rightarrow 0, \quad n \rightarrow \infty.
 \end{aligned}$$

Wir betrachten die Laplace-Transformierte von  $\bar{Z}$  und definieren

$$\begin{aligned}
 g(\lambda) &:= \mathbb{E} \exp(-\lambda \bar{Z}) \\
 h(s) &:= \mathbb{E} s^{\bar{K}} = \frac{(1-s)^\alpha + \alpha s - 1}{\alpha - 1},
 \end{aligned}$$

wobei die letzte Gleichung mittels (3.12) folgt. Aus (3.33) erhält man mit den bekannten Rechenregeln der Laplace-Transformation

$$\begin{aligned}
 g(\lambda) &= \int_0^1 h(g(\lambda u^{1/(\alpha-1)})) du \\
 &= \frac{1}{\alpha-1} \int_0^1 (1 - g(\lambda u^{1/(\alpha-1)}))^\alpha + \alpha g(\lambda u^{1/(\alpha-1)}) - 1 du.
 \end{aligned}$$

Setzt man  $v := \lambda u^{1/(\alpha-1)}$  so erhält man die Differentialgleichung

$$\lambda g'(\lambda) = g(\lambda) + (1 - g(\lambda))^\alpha - 1.$$

Mit  $u(\lambda) := 1 - g(\lambda)$  erkennt man eine Bernoulli Differentialgleichung

$$-\lambda u'(\lambda) = -u(\lambda) + u(\lambda)^\alpha,$$

deren Lösung

$$u(\lambda) = \frac{1}{(1 + c\lambda^{1-\alpha})^{\frac{1}{\alpha-1}}}$$

ist. Mit  $1 = \mathbb{E} \bar{Z} = g'(0) = -c^{-1/(\alpha-1)}$  gilt schließlich

$$g(\lambda) = 1 - \frac{\lambda}{(1 + \lambda^{\alpha-1})^{\frac{1}{\alpha-1}}}.$$

□

**Bemerkung.** Im Fall  $\sigma^2 < \infty$  erhält man im Vergleich zu (3.33) die Fixpunktgleichung

$$\mathcal{L}(\bar{Z}) = \mathcal{L}(U(\bar{Z}_1 + \bar{Z}_2)),$$

wobei  $U$  uniform auf  $[0, 1]$  verteilt ist und  $\bar{Z}_1, \bar{Z}_2$  unabhängige Kopien von  $\bar{Z}$  und unabhängig von  $U$  seien. Diese Gleichung beschreibt die charakteristische Eigenschaft der Exponentialverteilung. Man sieht dies leicht ein, wenn man die relative Lage der ersten beiden Punkte eines homogenen Poissonprozesses auf  $\mathbb{R}^+$  betrachtet.



# Literaturverzeichnis

- [1] ALDOUS, D. (1991). Asymptotic fringe distributions for general families of random trees. *Ann. Appl. Probab.* **1**, 228–266.
- [2] ALDOUS, D. (1991). The continuum random tree. II: An overview. *Stochastic analysis (Durham, 1990)*, London Math. Soc. Lecture Note Ser. 167, 23–70.
- [3] ALDOUS, D. (1993). The continuum random tree III. *Ann. Probab.* **21**, 248–289.
- [4] ALDOUS, D. UND PITMAN, J. (1998). Tree-valued Markov chains derived from Galton-Watson processes. *Ann. Inst. Henri Poincaré* **34**, 637–686.
- [5] ATHREYA, K. B. UND NEY, P. (1972). *Branching Processes*. Springer, New York.
- [6] BENNIES, J. UND KERSTING, G. (2000). A random walk approach to Galton-Watson trees. *J. Theor. Probab.* **13**, 777–803.
- [7] BERTOIN, J. UND DONEY, R. A. (1994). On conditioning a random walk to stay positive. *Ann. Probab.* **22**, 2152–2167.
- [8] BINGHAM, N. H., GOLDIE, C. M. UND TEUGELS, J. L. (1989). *Regular variation*. Paperback ed., Cambridge University Press.
- [9] CHAUVIN, B., ROUAULT, A. UND WAKOLBINGER, A. (1991). Growing conditioned trees. *Stoch. Proc. Appl.* **39**, 117–130.
- [10] DUQUESNE, T. UND LE GALL, J.-F. (2002). *Random trees, Lévy processes and spatial branching processes*. *Astérisque* **287**, Paris.

- 
- [11] DUQUESNE, T. (2003). *A limit theorem for the contour process of conditioned Galton-Watson trees*. *Ann. Probab.* **31**, no. 2, 996-1027.
  - [12] DWASS, M. (1969). The total progeny in a branching process and a related random walk. *J. Appl. Probab.* **6**, 682-686.
  - [13] FELLER, W. (1971). *Probability and its Applications*. 2nd ed., Wiley, New York.
  - [14] GEIGER, J. (1996). Size-biased and conditioned random splitting trees. *Stoch. Proc. Appl.* **65**, 187-207.
  - [15] GEIGER, J. (1999). Elementary new proofs of classical limit theorems for Galton-Watson processes. *J. Appl. Probab.* **36**, 301-309.
  - [16] GEIGER, J. (2000). A new proof of Yaglom's exponential limit law, *D. Gardy and A. Mokkadem, eds., Algorithms, Trees, Combinatorics and Probabilities, Trends in Mathematics*, Birkhäuser, 245-249.
  - [17] GEIGER, J. UND KERSTING, G. (1999). The Galton-Watson tree conditioned on its height. In: B. Grigelionis et al, eds., *Proceedings of the 7th Vilnius Conference 1998*, VSP, Utrecht, 277-286.
  - [18] GREVEN, A., KLENKE, A. UND WAKOLBINGER, A. (1999). The longtime behavior of branching random walk in a catalytic medium. *Electron. J. Probab.* **4**, no. 12, 1-80.
  - [19] HARRIS, T. (1963). *The Theory of Branching Processes*. Springer, Berlin.
  - [20] JOFFE, A. UND WAUGH, W. A. O'N. (1982). Exact distributions of kin numbers in Galton-Watson processes. *J. Appl. Probab.* **19**, 767-775.
  - [21] KALLENBERG, O. (1977). Stability of critical cluster fields. *Math. Nachr.* **77**, 7-43.
  - [22] KENNEDY, D. J. (1975). The Galton-Watson process conditioned on the total progeny. *J. Appl. Probab.* **12**, 800-806.
  - [23] KESTEN, H., NEY, P. UND SPITZER, F. (1966). The Galton-Watson process with mean one and finite variance. *Theory Prob. Appl.* **11**, 513-540.

- 
- [24] KOLCHIN, V. F. (1986). *Random Mappings*. Optimization Software, New York.
  - [25] LYONS, R., PEMANTLE, R. UND PERES, Y. (1995). Conceptual proofs of  $L \log L$  criteria for mean behavior of branching processes. *Ann. Probab.* **23**, 1125–1138.
  - [26] MEIR, A. UND MOON, J. W. (1978). On the altitude of nodes in random trees. *Can. J. Math.* **30**, 997–1015.
  - [27] NEININGER, R. UND RÜSCHENDORF, L. (2003). A general limit theorem for recursive algorithms and combinatorial structures. *Ann. Appl. Probab.*, accepted for publication.
  - [28] NEVEU, J. (1986). Arbres et processus de Galton-Watson. *Ann. Inst. Henri Poincaré* **22**, 199–207.
  - [29] PITMAN, J. (1998). Enumerations of trees and forests related to branching processes and random walks. *DIMACS Series in Discrete Mathematics and Theoretical Computer Science* **41**, 163–180.
  - [30] SLACK, R.S. 1967. A branching process with mean one and possibly infinite variance. *Z. Wahrscheinlichkeitstheorie verw. Geb.* **9**, 139–145
  - [31] TAKÁCS, L. (1993). The asymptotic distribution of the total heights of random rooted trees. *Acta Sci. Math.* **57**, 613–625.
  - [32] YAGLOM, A.M. (1947). Certain limit theorems of the theory of branching processes. *Doklady. Akad. Nauk. SSSR* **56**, 795–798.
  - [33] ZUBKOV, A.M. (1975). Limiting distributions of the distance the closest common ancestor. *Theory Prob. Appl.* **20**, 602–612.
  - [34] ZOLOTAREV, V.M. (1976). Approximation of the distributions of sums of independent random variables with values in infinite-dimensional spaces. *Theory Prob. Appl.* **21**, 721–737.





Lars Kauffmann  
Scheidswaldstraße 53  
60385 Frankfurt am Main

## Lebenslauf

Geboren am 29.01.1971 in Friedberg (Hessen) als Sohn von Christoph Kauffmann und Hannelore Kauffmann, geborene Dietz.

- |           |                                                                                                                                                                                                                                                                                                     |
|-----------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1977-1987 | Grundschule in Ober-Mörlen, Förderstufe in Nieder-Mörlen, Realschule in Bad Nauheim.                                                                                                                                                                                                                |
| 1987-1990 | Berufliches Gymnasium der Johann Philipp Reiss Schule in Friedberg, Leistungskurse Elektrotechnik und Mathematik, Allgemeine Hochschulreife.                                                                                                                                                        |
| 1990-1998 | Mathematikstudium mit Nebenfach Informatik an der Johann Wolfgang Goethe-Universität in Frankfurt am Main, Vordiplom im Herbst 1992, Diplomarbeit bei Prof. G. Kersting mit dem Titel " <i>Über quasistationäre Verteilungen von Geburts- und Todesprozessen</i> ", Diplom Mathematik im März 1998. |
| 1998-2003 | Wissenschaftlicher Mitarbeiter des Instituts für Stochastik und mathematische Informatik am Fachbereich Mathematik der Universität Frankfurt (bei Prof. H. Dinges). Bearbeitung des Promotionsthemas bei Prof. G. Kersting.                                                                         |